arXiv:1304.7965v1 [math.FA] 30 Apr 2013

Analysis of the convergence rate for the cyclic projection algorithm applied to semi-algebraic convex sets

Jonathan M. Borwein[*], Guoyin Li [†], and Liangjin Yao[‡]

April 29, 2013

Abstract

In this paper, we study the rate of convergence of the cyclic projection algorithm applied to finitely many semi-algebraic convex sets. We establish an explicit convergence rate estimate which relies on the maximum degree of the polynomials that generate the semi-algebraic convex sets and the dimension of the underlying space. We achieve our results by exploiting the algebraic structure of the semi-algebraic convex sets.

2010 Mathematics Subject Classification:
Primary 41A25, 90C25; Secondary 41A50, 90C31

Keywords: Cyclic projection algorithm, convex polynomial, distance function, Fejér monotone sequence, Hölderian regularity, Łojasiewicz's inequality, projector operator, semi-algebraic convex set, von Neumann alternating projection method.

1 Introduction

A very common problem in diverse areas of mathematics and engineering consists of trying to find a point in the intersection of closed convex sets C_i, $i = 1, \dots, m$. This problem is often referred to as the convex feasibility problem. One popular method for solving the convex feasibility problem is the so-called cyclic projection algorithm. Mathematically, the cyclic projection algorithm is formulated as follows. Given finite many closed convex sets

[*]CARMA, University of Newcastle, Newcastle, New South Wales 2308, Australia. E-mail: jonathan.borwein@newcastle.edu.au. Laureate Professor at the University of Newcastle and Distinguished Professor at King Abdul-Aziz University, Jeddah.

[†]Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia. E-mail: g.li@unsw.edu.au.

[‡]CARMA, University of Newcastle, Newcastle, New South Wales 2308, Australia. E-mail: liangjin.yao@newcastle.edu.au.

$C_1, C_2, \cdots, C_m$ in $\mathbb{R}^n$ with $\bigcap_{i=1}^m C_i \neq \varnothing$. Let $x_0 \in \mathbb{R}^n$ and $P_i := P_{C_i}, i = 1, 2, \cdots, m$, where P_{C_i} denotes the Euclidean projection to the set C_i. The sequence of *cyclic projections*, $(x_k)_{k\in\mathbb{N}}$, is defined by

$$x_1 := P_1 x_0, x_2 := P_2 x_1, \cdots, x_m := P_m x_{m-1}, x_{m+1} := P_1 x_m \ldots \tag{1.1}$$

Bregman showed that the sequence $(x_k)_{k\in\mathbb{N}}$ generated by the cyclic projection algorithm, converges to a point in C (see [18] and [9, 5]). When $m = 2$, the cyclic projection method reduces to the well known von Neumann alternating projection method (APM) (see [44] and also [6, 8, 10, 11]). The cyclic projection method has attracted many interests recently due to its simplicity and numerous applications to diverse areas such as engineering and the physical sciences, see [18, 26, 15, 3, 22, 8, 9, 5, 4, 23, 24] and the references therein.

In this paper, we focus on the case where each set C_i is a *semi-algebraic convex set* in the sense that there exist $\gamma_i \in \mathbb{N}$ and convex polynomial functions, $g_{i\,j}, j = 1, \ldots, \gamma_i$ such that

$$C_i = \{x \in \mathbb{R}^n \mid g_{i\,j}(x) \leq 0, j = 1, \cdots, \gamma_i\}.$$

We then provide an explicit rate for the cyclic projection algorithm involving finitely many semi-algebraic convex sets *without any regularity conditions*. More precisely, let C_i be semi-algebraic convex sets generated by polynomials in $\mathbb{R}^n$ with degree at most d, $d \in \mathbb{N}\backslash\{1\}$. We show that the sequence of cyclic projections $(x_k)_{k\in\mathbb{N}}$ (1.1) converges (at least) at the rate of $\frac{1}{k^\rho}$, where $\rho := \frac{1}{\min\left\{(2d-1)^n-1,\, 2\beta(n-1)d^n-2\right\}}$ and $\beta(s)$ denotes the *central binomial coefficient* with respect to s which is given by $\binom{s}{[s/2]}$. [1]

The remainder of this paper is organized as follows. In Section 2, we collect notations and auxiliary results for future reference and for the reader's convenience. In Section 3, we give a Hölderian regularity result for finitely many semi-algebraic convex sets. The proof of our main result (Theorem 4.2) forms the bulk of Section 4. In Section 5, we explore various concrete examples. Finally, we end the paper with some conclusions and open questions.

2 Preliminaries and auxiliary results

We assume throughout that $\mathbb{R}^n$ is a Euclidean space with the norm $\|\cdot\|$ and the inner product $\langle\cdot,\cdot\rangle$, where $n \in \mathbb{N} := \{1, 2, 3, \cdots\}$. We reserve $d \in \mathbb{N}$. We denote by $\mathbb{B}(x, \varepsilon) := \{y \in \mathbb{R}^n \mid \|y - x\| < \varepsilon\}$. We adopt standard notation used in these books [10, 14, 19, 20, 41, 42, 45].

Given a subset C of $\mathbb{R}^n$, $\operatorname{int} C$ is the *interior* of C, $\operatorname{bd} C$ is the *boundary* of C, $\operatorname{aff} C$ is the *affine hull* of C and $\overline{C}$ is the norm *closure* of C. We set $C^\perp := \{x^* \in \mathbb{R}^n \mid (\forall c \in$

[1] Here, $[a]$ denotes the integer part of a

$C) \langle x^*, c\rangle = 0\}$. The *distance function* to the set C, written as $\operatorname{dist}(\cdot, C)$, is defined by $x \mapsto \inf_{c\in C} \|x - c\|$. The *projector operator* to the set C, denoted by P_C, is defined by

$$P_C(x) := \{c \in C \mid \|y - c\| = \operatorname{dist}(x, C)\}, \quad \forall x \in \mathbb{R}^n.$$

Let $D \subseteq \mathbb{R}^n$. The distance of two sets: C and D, is $\operatorname{dist}(C, D) := \inf_{c\in C, d\in d} \|c-d\|$. Given $f\colon X \to\ ]-\infty, +\infty]$, we set $\operatorname{dom} f := f^{-1}(\mathbb{R})$. We say f is *proper* if $\operatorname{dom} f \neq \varnothing$. Let f be a proper function on $\mathbb{R}^n$. Its associated recession function f^∞ is defined by

$$f^\infty(v) := \liminf_{t\to\infty,\, v'\to v} \frac{f(tv')}{t} \quad \text{for all } v \in \mathbb{R}^n.$$

If f is further assumed to be lower semicontinuous and convex, one has (see [1, Proposition 2.5.2])

$$f^\infty(v) = \lim_{t\to\infty} \frac{f(x + tv) - f(x)}{t} = \sup_{t>0} \frac{f(x + tv) - f(x)}{t} \quad \text{for all } x \in \operatorname{dom} f. \tag{2.1}$$

Recall that $f\colon \mathbb{R}^n \to \mathbb{R}$ is a *polynomial* if there exists a number $r \in \mathbb{N}$ such that

$$f(x) := \sum_{0\le|\alpha|\le r} \lambda_\alpha x^\alpha,$$

where $\lambda_\alpha \in \mathbb{R}$, $x = (x_1, \cdots, x_n)$, $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, $\alpha_i \in \mathbb{N} \cup \{0\}$, and $|\alpha| := \sum_{j=1}^n \alpha_j$. The corresponding constant r is called the *degree* of f.

Next let us recall a useful property of polynomial functions.

Fact 2.1 (See [2, Remark 4]) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be polynomial, and $\{x_1, x_2\} \subseteq \mathbb{R}^n$. If f is constant on $D := [x_1, x_2]$, then f is constant on* $\operatorname{aff} D$.

Following [17], a set $D \subseteq \mathbb{R}^n$ is said to be *semi-algebraic* if

$$D := \bigcup_{i=1}^{l} \bigcap_{j=1}^{s} \{x \in \mathbb{R}^n \mid f_{ij}(x) = 0, h_{ij}(x) \le 0\}$$

for some integers l, s and some polynomial functions f_{ij}, h_{ij} on $\mathbb{R}^n$ ($1 \le i \le l$, $1 \le j \le s$). Moreover, a function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *semi-algebraic* if its graph $\operatorname{gph} f := \{(x, f(x)) \mid x \in \mathbb{R}^n\}$ is semi-algebraic.

Recall that a set $C \subseteq \mathbb{R}^n$ is called semi-algebraic convex if there exist $\gamma \in \mathbb{N}$ and convex polynomial functions, $g_j, j = 1, \dots, \gamma$ such that $C = \{x \in \mathbb{R}^n \mid g_j(x) \le 0, j = 1, \cdots, \gamma\}$. Clearly, any semi-algebraic convex set is convex and semi-algebraic. However, the following example shows that a convex and semi-algebraic set need not to be a semi-algebraic convex set.

Example 2.2 Consider a set $A := \{(x_1, x_2) \in \mathbb{R}^2 \mid 1 - x_1x_2 \le 0, -x_1 \le 0, -x_2 \le 0\}$. Clearly, A is convex and semi-algebraic while the polynomial $(x_1, x_2) \mapsto 1 - x_1x_2$ is not

convex. We now show that A is not a semi-algebraic convex set, i.e., it cannot be written as $\{x : g_i(x) \le 0, i = 1, \cdots, l\}$ for some convex polynomials g_i, $i = 1, \cdots, l$, $l \in \mathbb{N}$. To see this, we proceed by the method of contradiction. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x_1, x_2) := x_1$. Clearly $\inf_{x=(x_1,x_2)\in A} f(x) = 0$. Then, then by Fact 2.13, f should attain its minimum on A. This makes contradiction, and so, justify the claim. $\diamondsuit$

We now summarize below some basic properties of semi-algebraic sets and semi-algebraic functions. These properties will be useful for our discussion later.

Fact 2.3 *The following statements hold (the properties (P1) and (P4) are direct from the definitions).*

(P1) Any polynomial is a semi-algebraic function.

(P2) (See [17, Proposition 2.2.8].) *Let D be a semi-algebraic set. Then* $\mathrm{dist}(\cdot, D)$ *is a semi-algebraic function.*

(P3) (See [17, Proposition 2.2.6].) *If f, g are semi-algebraic functions on $\mathbb{R}^n$ and $\lambda \in \mathbb{R}$ then $f + g$, λf, $\max\{f, g\}$, fg are semi-algebraic.*

(P4) If f_i are polynomials, $i = 1, \ldots, m$, and $\lambda \in \mathbb{R}$, then the sets $\{x \mid f_i(x) = \lambda, i = 1, \ldots, m\}$, $\{x \mid f_i(x) \le \lambda, i = 1, \ldots, m\}$ are semi-algebraic sets.

*(P5) (*Łojasiewicz's inequality*) (See [17, Corollary 2.6.7].)* If ϕ, ψ are two continuous semi-algebraic functions on compact semi-algebraic set $K \subseteq \mathbb{R}^n$ such that $\emptyset \neq \phi^{-1}(0) \subseteq \psi^{-1}(0)$ then there exist constants $c > 0$ and $\tau \in (0, 1]$ such that

$$|\psi(x)| \le c|\phi(x)|^\tau \quad \text{for all } x \in K.$$

Remark 2.4 As pointed out by [38], the corresponding exponent τ in the Łojasiewicz's inequality (P5) is hard to determine and is typically unknown. $\diamondsuit$

Remark 2.5 Let g_i, $i = 1, \ldots, m$ be polynomial on $\mathbb{R}^n$ and let $S := \{x \in \mathbb{R}^n \mid g_i(x) \le 0\}$. Let $\bar{x} \in S$. Then, (P2) and (P4) imply that $\phi = \max_{1\le i\le m}[g_i]_+$ where $[g_i]_+ := \max\{g_i(\cdot), 0\}$ and $\psi = \mathrm{dist}(\cdot, S)$ are semi-algebraic functions. Applying (P5) it follows that there exist $c, \varepsilon > 0$ and $\tau \in (0, 1]$ such that

$$\mathrm{dist}(x, S) \le c \max_{1\le i\le m} [g_i(x)]_+^\tau \quad \text{for all } x \in \mathbb{B}(\bar{x}, \varepsilon). \tag{2.2}$$

$\diamondsuit$

As we have explained in Remark 2.4, the exponent τ in (2.2) is hard to determine and is typically unknown. However, there are some special cases where we can provide some effective estimates on the exponent τ: To formulate these results, we first introduce a notation below.

Definition 2.6 *Define*

$$\kappa(n,d) := (d-1)^n + 1. \tag{2.3}$$

We now present various results which explain that the exponent τ in (2.2) can be effectively estimated in the case when g_i has some special structure.

Fact 2.7 (Gwoździewicz) (See [25, Theorem 3].) *Let g be a polynomial on $\mathbb{R}^n$ with degree no larger than d. Suppose that $g(0) = 0$ and there exists $\varepsilon_0 > 0$ such that $g(x) > 0$ for all $x \in \mathbb{B}(0, \varepsilon_0)\backslash\{0\}$. Then there exist constants $c, \varepsilon > 0$ such that*

$$\|x\| \leq c\, g(x)^{\frac{1}{\kappa(n,d)}}, \quad \forall x \in \mathbb{B}(0, \varepsilon). \tag{2.4}$$

We denote by $\beta(s)$ the *central binomial coefficient* with respect to an integer s: $\binom{s}{[s/2]}$ (with $\binom{0}{0} = 1$)[29].

Fact 2.8 (Kollár) (See [29, Theorem 3(i)].) *Let g_i be polynomials on $\mathbb{R}^n$ with degree $\leq d$ for every $i = 1, \cdots, m$. Let $g(x) := \max_{1\leq i\leq m} g_i(x)$. Suppose that there exists $\varepsilon_0 > 0$ such that $g(x) > 0$ for all $x \in \mathbb{B}(0, \varepsilon_0)\backslash\{0\}$. Then there exist constants $c, \varepsilon > 0$ such that*

$$\|x\| \leq c\, g(x)^{\frac{1}{\beta(n-1)d^n}}, \quad \forall x \in \mathbb{B}(0, \varepsilon).$$

Fact 2.9 (See [32, Theorem 4.2].) *Let g be a convex polynomial on $\mathbb{R}^n$ with degree at most d. Let $S := \{x \mid g(x) \leq 0\}$ and $\bar{x} \in S$. Then, g has a Hölder type local error bound with exponent $\kappa(n,d)^{-1}$, i.e., there exist constants $c, \varepsilon > 0$ such that*

$$\operatorname{dist}(x, S) \leq c\, [g(x)]_+^{\frac{1}{\kappa(n,d)}} \quad \textit{for all } x \in \mathbb{B}(\bar{x}, \varepsilon).$$

Fact 2.10 (See [32, Theorem 4.1].) *Let g be a convex polynomial on $\mathbb{R}^n$. Let $S := \{x \mid g(x) \leq 0\}$. Suppose that there exists $x_0 \in \mathbb{R}^n$ such that $g(x_0) < 0$. Then, g has a Lipschitz type global error bound, i.e., there exists a constant $c > 0$ such that*

$$\operatorname{dist}(x, S) \leq c\, [g(x)]_+ \quad \textit{for all } x \in \mathbb{R}^n.$$

Corollary 2.11 *Let g_i be convex polynomials on $\mathbb{R}^n$ for every $i = 1, \cdots, m$. Let $S := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \cdots, m\}$. Suppose that there exists $x_0 \in \mathbb{R}^n$ such that $g_i(x_0) < 0$ for every $i = 1, \cdots, m$. Then, for every compact subset K of $\mathbb{R}^n$, there exists $c > 0$ such that*

$$\operatorname{dist}(x, S) \leq c \max_{1\leq i\leq m} [g_i(x)]_+ \quad \textit{for all } x \in K.$$

Proof. Let K be a compact subset of $\mathbb{R}^n$. We will prove it by the induction on m. When $m = 1$, the conclusion follows by applying Fact 2.10 directly. Suppose the statement holds when $m = p - 1, p \in \mathbb{N}\backslash\{1\}$. Then, there exists a constant $c_0 > 0$ such that

$$\operatorname{dist}(x, S_1) \leq c_0 \max_{1\leq i\leq p-1} [g_i(x)]_+ \quad \text{for all } x \in K, \tag{2.5}$$

where $S_1 := \{x \in \mathbb{R}^n \ \mid g_i(x) \le 0, i = 1, \cdots, p-1\}$

Now we consider the case of $m = p$. Set $S_2 := \{x \in \mathbb{R}^n \ \mid g_p(x) \le 0\}$. Then $x_0 \in \operatorname{int} S_1 \cap \operatorname{int} S_2$. By [6, Corollary 4.5], there exists $\gamma > 0$ such that for every $x \in K$

$$
\begin{aligned}
&\operatorname{dist}(x, S) = \operatorname{dist}(x, S_1 \cap S_2) \le \gamma \max\big\{\operatorname{dist}(x, S_1), \operatorname{dist}(x, S_2)\big\} \\
&\le \gamma \max\Big\{c_0 \max_{1 \le i \le p-1} [g_i(x)]_+, \operatorname{dist}(x, S_2)\Big\} \quad \text{(by (2.5))} \\
&\le \gamma \max\Big\{c_0 \max_{1 \le i \le p-1} [g_i(x)]_+, c_1 [g_m(x)]_+\Big\} \quad (\exists c_1 > 0 \text{ by Fact 2.10}) \\
&\le c \max_{1 \le i \le p} [g_i(x)]_+,
\end{aligned}
$$

where $c := \max\{\gamma c_0, \gamma c_1\}$.

Thus, the conclusion follows by the method of induction. □

The following example will show us that the conclusion of Corollary 2.11 can fail if we allow K to be noncompact.

Example 2.12 (Shironin) (See [32, Example 4.1] or [43].) Let $g_1, g_2 : \mathbb{R}^4 \to \mathbb{R}$ be defined by

$$
\begin{aligned}
g_1(x_1, x_2, x_3, x_4) &:= x_1, \\
g_2(x_1, x_2, x_3, x_4) &:= x_1^{16} + x_2^8 + x_3^6 + x_1 x_2^3 x_3^3 + x_1^2 x_2^4 x_3^2 + x_2^2 x_3^4 + x_1^4 x_3^4 \\
&\quad + x_1^4 x_2^6 + x_1^2 x_2^6 + x_1^2 + x_2^2 + x_3^2 - x_4.
\end{aligned}
$$

Then g_1,g_2 are convex polynomials and

$$
g_1(-k, 0, 0, k^{16} + k^2 + k) = g_2(-k, 0, 0, k^{16} + k^2 + k) = -k < 0, \quad \forall k \in \mathbb{N}.
$$

On the other hand, as shown in [43], there exists a sequence $(x_k)_{k \in \mathbb{N}}$ in $\mathbb{R}^4$ such that

$$
\max\big\{[g_1(x_k)]_+, [g_2(x_k)]_+\big\} \le 1, \forall k \quad \text{but} \quad \operatorname{dist}(x_k, S) \longrightarrow +\infty,
$$

where $S := \{x \in \mathbb{R}^4 \mid g_1(x) \le 0, g_2(x) \le 0\}$. ◊

We now summarize some basic properties of convex polynomials that will be used later. The first property is a Frank-Wolfe type result for convex polynomial optimization problems while the second one is a directional constant property for a convex polynomial.

Fact 2.13 (Belousov) (See [13, Theorem 13, Section 4, Chapter II] or [16, Theorem 3] and [39].) *Let f be a convex polynomial on $\mathbb{R}^n$. Consider a set $D := \{x \mid g_i(x) \le 0, i = 1, \cdots, m\}$, where each g_i, $i = 1, \cdots, m$, is a convex polynomial on $\mathbb{R}^n$. Suppose that $\inf_{x \in D} f(x) > -\infty$. Then f attains its minimum on D.*

Fact 2.14 (See [1, Propsition 3.2.1].) *Let f be a convex polynomial on $\mathbb{R}^n$ and $v \in \mathbb{R}^n$. Assume that $f^\infty(v) = 0$. Then $f(x + tv) = f(x)$ for all $t \in \mathbb{R}$ and for all $x \in \mathbb{R}^n$.*

From now on we assume that

$$
\boxed{
\begin{aligned}
& m \in \mathbb{N}, \gamma_i \in \mathbb{N},\, i = 1, \ldots, m \\
& g_{i,1}, g_{i,2}, \cdots, g_{i,\gamma_i} \text{ are convex polynomials on } \mathbb{R}^n,\, i = 1, \ldots, m \\
& C_i := \Big\{ x \in \mathbb{R}^n \mid g_{i,1}(x) \le 0, g_{i,2}(x) \le 0, \cdots, g_{i,\gamma_i}(x) \le 0 \Big\},\, i = 1, \ldots, m \\
& P_i := P_{C_i}, \quad \forall i = 1, 2, \cdots, m \\
& C := \bigcap_{i=1}^m C_i \neq \varnothing.
\end{aligned}
}
$$

Fact 2.15 (Bauschke and Borwein) (See [7, Lemma 2.2 and Theorem 4.8], or [6, Fact 1.1(iii) and Fact 1.2(ii)].) *Let A, B be nonempty convex subsets of $\mathbb{R}^n$ such that $A - B$ is closed. Let $b_0 \in X$ and $(a_k)_{k\in\mathbb{N}}, (b_k)_{k\in\mathbb{N}}$ be defined as below:*

$$
\begin{cases}
a_{k+1} & := P_A b_k \\
b_{k+1} & := P_B a_{k+1}.
\end{cases}
$$

Let $v := P_{A-B}0$.

(i) $\|v\| = \operatorname{dist}(A, B)$ *and* $a_k \longrightarrow a, \quad b_k \longrightarrow a + v$.

(ii) $P_B x = P_{B\cap(A+v)} x = x + v,\ \forall x \in A \cap (B - v)$ *and* $P_A y = P_{A\cap(B-v)} y = y - v,\ \forall y \in B \cap (A + v)$

Definition 2.16 *Let A be a nonempty convex subset of $\mathbb{R}^n$. We say the sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ is* Fejér monotone *with respect to A if*

$$\|x_{k+1} - a\| \le \|x_k - a\|, \quad \forall k \in \mathbb{N},\, a \in A.$$

Fact 2.17 (Bauschke and Borwein) (See [6, Theorem 3.3(iv)].) *Let A be a nonempty closed convex subset of $\mathbb{R}^n$ and $(x_k)_{k\in\mathbb{N}}$ be Fejér monotone with respect to A, and $x_k \longrightarrow x \in A$. Then $\|x_n - x\| \le 2\operatorname{dist}(x_n, A)$.*

Fact 2.18 (Bregman) (See [18].) *Let $x_0 \in \mathbb{R}^n$. The sequence of cyclic projections, $(x_k)_{k\in\mathbb{N}}$, defined by*

$$x_1 := P_1 x_0, x_2 := P_2 x_1, \cdots, x_m := P_m x_{m-1}, x_{m+1} := P_1 x_m \ldots \tag{2.6}$$

converges to a point in C.

3 Hölderian regularity for semi-algebraic convex sets

In this section, we will derive the Hölderian regularity for semi-algebraic convex sets and provide an effective estimate of the exponent in the regularity results. This result plays an important role in quantifying the convergence speed of the cyclic projection methods.

To do this, we first establish an error bound result which estimates the distance of a point to a semi-algebraic convex set S in terms of the polynomials that defines S. More explicitly, we derive the explicit exponent $\tau > 0$ such that there exist $c, \varepsilon > 0$,

$$\mathrm{dist}(x, S) \le c\bigg(\max_{1 \le i \le m} [g_i(x)]_+\bigg)^{\tau} \quad \textit{whenever} \quad \|x - \overline{x}\| \le \epsilon,$$

where $S := \{x \in \mathbb{R}^n \mid g_i(x) \le 0, i = 1, \cdots, m\}$. We note that the error bound property plays an important role in convergence analysis of many algorithms for optimization problems as well as variational inequality problem [33, 37, 40], and the exponent τ in the error bound property has a close relationship with the convergence rate of the algorithm. However, existing results such as the powerful Łojasiewicz's inequality do not provide any insight on how to explicitly estimate the exponent τ.

Before we proceed, let us use a simple example to illustrate that the exponent τ can be related to the maximum degree of the polynomials defined the semi-algebraic convex set and the dimension of the underlying space. This example is partially inspired by [29, Example 1].

Example 3.1 Let d be an even number. Consider convex polynomials g_i, $i = 1, \ldots, n$ on $\mathbb{R}^n$ given by $g_1(x) := x_1^d$ and g_i, $i = 2, \cdots, n$ given by $g_i(x) := x_i^d - x_{i-1}$, $i = 2, \cdots, n$. Then, direct verification gives us that $S := \{x \in \mathbb{R}^n \mid g_i(x) \le 0, i = 1, \cdots, n\} = \{0\}$, and so, $\mathrm{dist}(x, S) = \|x\|$. In this case, consider $x(t) = (t^{d^{n-1}}, t^{d^{n-2}}, \cdots, t) \in \mathbb{R}^n$, $t \in (0, 1)$. Then $\mathrm{dist}(x(t), S) = O(t)$ and $\max_{1 \le i \le m}[g_i(x(t))]_+ = t^{d^n}$. Therefore, we see that if there exist $c, \epsilon > 0$ and $\tau > 0$ such that

$$\mathrm{dist}(x, S) \le c\bigg(\max_{1 \le i \le m} [g_i(x)]_+\bigg)^{\tau} \quad \textit{whenever} \quad \|x\| \le \epsilon,$$

then, $\tau \le \frac{1}{d^n}$. Thus, we see that the exponent τ is related to the maximum degree of the polynomials defined the semi-algebraic convex set and the dimension of the underlying space. $\diamondsuit$

We now introduce a decomposition of the index set.

Definition 3.2 *For convex polynomials $g_1, \ldots, g_m$ on $\mathbb{R}^n$ with $S = \{x \in \mathbb{R}^n \mid g_i(x) \le 0, i = 1, \ldots, m\}$, the index set $\{1, \ldots, m\}$ can be decomposed as $J_1 \cup J_2$ with $J_1 \cap J_2 = \emptyset$ where*

$$J_1 := \big\{i \in \{1, \ldots, m\} \mid g_i(S) \equiv 0\big\} \quad \textit{and} \quad J_2 := \{1, \ldots, m\} \backslash J_1. \tag{3.1}$$

Now we come to our key technical result.

Theorem 3.3 (Local error bounds for convex polynomial systems) *Let g_i be convex polynomials on $\mathbb{R}^n$ with degree at most d for every $i = 1, \cdots, m$. Let $S := \{x \in \mathbb{R}^n \mid g_i(x) \le 0, i = 1, \cdots, m\}$ and $\bar{x} \in S$. Then there exist $c, \varepsilon > 0$ such that*

$$\operatorname{dist}(x, S) \le c\Big(\max_{i\in J_2}[g_i(x)]_+ + \big(\max_{i\in J_1}[g_i(x)]_+\big)^{\tau}\Big) \quad \textit{whenever} \quad \|x - \overline{x}\| \le \varepsilon,$$

where $[a]_+ := \max\{a, 0\}$, $\tau := \max\left\{\frac{2}{\kappa(n,2d)}, \frac{1}{\beta(n-1)d^n}\right\}$, $\kappa(n, 2d) := (2d-1)^n + 1$, $\beta(n-1)$ is the central binomial coefficient with respect to $n-1$ which is given by $\binom{n-1}{[(n-1)/2]}$, and J_1, J_2 are defined as in (3.1).

Proof. We prove the desire conclusion by induction on the number of the polynomials m.

[**Trivial Case**] Suppose that $m = 1$. Then $J_1 = \{1\}$ or $J_1 = \emptyset$. If $J_1 = \{1\}$, then the conclusion follows by Fact 2.9 since $\max\left\{\frac{2}{\kappa(n,2d)}, \frac{1}{\beta(n-1)d^n}\right\} \le \frac{1}{\kappa(n,d)}$. If $J_1 = \emptyset$, then there exists x_0 such that $g_1(x_0) < 0$. In this case, the conclusion follows by Fact 2.10.

[**Reduction to the active cases**] Let us suppose that the conclusion is true for $m \le p-1$, $p \in \mathbb{N}$, and look at the case for $m = p$. If $J_1 \ne \{1, \dots, m\}$, then $\{1, \dots, m\}\backslash J_1 \ne \emptyset$. Let $i_0 \notin J_1$. Then there exists $x_0 \in S$ such that $g_{i_0}(x_0) < 0$. Set $J_0 := \big\{i \in \{1, 2, \dots, m\} \mid g_i(x_0) < 0\big\}$. Then $i_0 \in J_0 \subseteq J_2$ and $J_0 \cap J_1 = \varnothing$. Let A, B be defined by

$$A := \big\{x \in \mathbb{R}^n \mid g_i(x) \le 0, \quad \forall i \in J_0\big\}$$
$$B := \big\{x \in \mathbb{R}^n \mid g_i(x) \le 0, \quad \forall i \in \widetilde{J}\big\},$$

where $\widetilde{J} := \{1, 2, \dots, m\}\backslash J_0$. Thus, $J_1 \subseteq \widetilde{J}$. Then we have $x_0 \in \operatorname{int} A \cap B$ and $S = A \cap B$. Since $S \subseteq B$, we have

$$\widetilde{J_1} := \big\{i \in \{1, 2, \dots, m\}\backslash J_0 \mid g_i(B) \equiv 0\big\} \subseteq J_1. \tag{3.2}$$

Since $\operatorname{int} A \cap B \ne \emptyset$, [6, Corollary 4.5] implies that for every compact set K there exist $\gamma, \delta > 0$ such that

$$\operatorname{dist}(x, S) = \operatorname{dist}(x, A \cap B) \le \gamma \max\{\operatorname{dist}(x, A), \operatorname{dist}(x, B)\} \quad \text{for all } x \in K. \tag{3.3}$$

By Corollary 2.11, there exists $c_1 > 0$ such that

$$\operatorname{dist}(x, A) \le c_1 \max_{i\in J_0}[g_i(x)]_+, \quad \forall x \in K. \tag{3.4}$$

From the induction hypothesis and (3.2), we see that there exist $\varepsilon > 0$ and $c_2 > 0$ such that for every $\|x - \bar{x}\| \le \varepsilon$, $\max_{1\le i\le m}[g_i(x)]_+ \le 1$ and

$$\operatorname{dist}(x, B)$$
$$\le c_2\Big(\max_{i\in \widetilde{J}\backslash \widetilde{J1}}[g_i(x)]_+ + \big(\max_{i\in \widetilde{J}_1}[g_i(x)]_+\big)^{\tau}\Big)$$

$$
\begin{aligned}
&= c_2\Big(\max\big\{ \max_{i\in \widetilde{J}\setminus J_1} [g_i(x)]_+, \max_{i\in J_1\setminus \widetilde{J1}} [g_i(x)]_+ \big\} + \big(\max_{i\in \widetilde{J_1}} [g_i(x)]_+\big)^\tau \Big) \quad (\text{since } \widetilde{J}_1 \subseteq J_1 \subseteq \widetilde{J}) \\
&\le c_2\Big(\max_{i\in \widetilde{J}\setminus J_1} [g_i(x)]_+ + \max_{i\in J_1\setminus \widetilde{J1}} [g_i(x)]_+ + \big(\max_{i\in \widetilde{J_1}} [g_i(x)]_+\big)^\tau \Big) \\
&\le c_2\Big(\max_{i\in \widetilde{J}\setminus J_1} [g_i(x)]_+ + \big(\max_{i\in J_1\setminus \widetilde{J1}} [g_i(x)]_+\big)^\tau + \big(\max_{i\in \widetilde{J_1}} [g_i(x)]_+\big)^\tau \Big) \\
&\le c_2\Big(\max_{i\in \widetilde{J}\setminus J_1} [g_i(x)]_+ + 2\big(\max_{i\in J_1} [g_i(x)]_+\big)^\tau \Big) \\
&\le 2c_2\Big(\max_{i\in \widetilde{J}\setminus J_1} [g_i(x)]_+ + \big(\max_{i\in J_1} [g_i(x)]_+\big)^\tau \Big).
\end{aligned}
$$

Thus the conclusion follows in this case by combining (3.3) and (3.4).

From now on, we may assume that $J_1 = \{1, \ldots, m\}$, that is,

$$\big\{x \mid g_i(x) \le 0, i = 1, \cdots, m\big\} = \big\{x \mid g_i(x) = 0, i = 1, \cdots, m\big\}.$$

This implies that $\inf_{x\in\mathbb{R}^n} \max_{1\le i\le m}\{g_i(x)\} = 0$. Then,

$$0_{\mathbb{R}^m} \notin \big\{(g_1(x), \cdots, g_m(x)) \mid x \in \mathbb{R}^n\big\} + \operatorname{int}\mathbb{R}^m_+.$$

Hence the convex separation theorem gives us that there exist $\alpha_i \ge 0$ with $\sum_{i=1}^m \alpha_i = 1$ such that $\sum_{i=1}^m \alpha_i g_i(x) \ge 0$ for all $x \in \mathbb{R}^n$. Denote $I := \{i \mid \alpha_i > 0\} \ne \emptyset$. Then, we have $\sum_{i\in I} \alpha_i = 1$ and

$$\sum_{i\in I} \alpha_i g_i(x) \ge 0 \quad \text{for all } x \in \mathbb{R}^n. \tag{3.5}$$

[Decompose the underlying space into sum of two subspaces M and $M^\perp$] Consider $D := \{x \in \mathbb{R}^n \mid g_i(x) \le 0, i \in I\}$. Clearly, D is a convex set and $\bar{x} \in D$. Moreover, for any $v \in D$, (3.5) implies that

$$g_i(v) = 0, \quad \forall i \in I. \tag{3.6}$$

In other words, g_i takes constant value 0 on D. Then, Fact 2.1 implies that D is either a singleton or an affine set with dimension larger than one.

Let $M := D - \bar{x}$. Then M is a subspace. We may decompose $\mathbb{R}^n = M + M^\perp$. Denote $\dim M = k$ $(k \le n)$.

We now see that

$$\sum_{i\in I} g_i^2(x) > 0 \text{ for all } x - \bar{x} \in M^\perp\setminus\{0\}. \tag{3.7}$$

Otherwise, there exists $x_0 \in \mathbb{R}^n$ such that $x_0 - \bar{x} \in M^\perp$ and $g_i(x_0) = 0$ for all $i \in I$. This shows that $x_0 \in D$. Thus $x_0 - \bar{x} \in M$ and hence $x_0 - \bar{x} \in M \cap M^\perp$. This contradicts to the fact that $x_0 - \bar{x} \ne 0$.

Similarly, we have

$$\max_{i\in I} g_i(x) > 0 \text{ for all } x-\bar{x} \in M^{\perp}\backslash\{0\}. \tag{3.8}$$

[**Distance estimation on** $M^{\perp}$] Now we first show that there exist $\varepsilon_0, \gamma_0 > 0$ such that

$$\|x-\bar{x}\| \le \gamma_0 \Big(\sum_{i\in I} g_i^2(x)\Big)^{\frac{1}{\kappa(n-k,2d)}}, \quad \forall x-\bar{x} \in M^{\perp}\cap \mathbb{B}(0,\varepsilon_0). \tag{3.9}$$

Since $\dim M^{\perp} = n-k$, there exists an $n\times(n-k)$ matrix Q_0 with the rank $n-k$ such that $Q_0(\mathbb{R}^{n-k}) = M^{\perp}$. Then Q_0 is a bijective operator from $\mathbb{R}^{n-k}$ to $M^{\perp}$. Then (3.7) shows that

$$\sum_{i\in I} g_i^2(\bar{x}+Q_0 b) > 0, \quad \forall b \in \mathbb{R}^{n-k}\backslash\{0\}. \tag{3.10}$$

Define $h:\mathbb{R}^{n-k}\to\mathbb{R}$ by $h(b) := \sum_{i\in I} g_i^2(\bar{x}+Q_0 b)$. Then $h(0) = \sum_{i\in I} g_i^2(\bar{x}) = 0$ by (3.6). Thus by (3.10) and Fact 2.7, there exist $\varepsilon_1, \gamma_1 > 0$ such that for all

$$\|b\| \le \gamma_1 h(b)^{\frac{1}{\kappa(n-k,2d)}} = \gamma_1 \Big(\sum_{i\in I} g_i^2(\bar{x}+Q_0 b)\Big)^{\frac{1}{\kappa(n-k,2d)}}, \quad \forall \|b\| \le \varepsilon_1.$$

Setting $x := \bar{x}+Q_0 b$, it follows that

$$\begin{aligned} \|x-\bar{x}\| &= \|Q_0\big(Q_0^{-1}(x-\bar{x})\big)\| \le \|Q_0\|\cdot\|Q_0^{-1}(x-\bar{x})\| \\ &\le \|Q_0\|\gamma_1\Big(\sum_{i\in I} g_i^2(x)\Big)^{\frac{1}{\kappa(n-k,2d)}}, \quad \forall \|Q_0^{-1}(x-\bar{x})\| \le \varepsilon_1. \end{aligned}$$

Then, there exist $\varepsilon_1, \gamma_1 > 0$ such that

$$\|x-\bar{x}\| \le \gamma_0 \Big(\sum_{i\in I} g_i^2(x)\Big)^{\frac{1}{\kappa(n-k,2d)}}, \quad \forall x-\bar{x} \in M^{\perp}\cap \mathbb{B}(0,\varepsilon_0).$$

Hence (3.9) holds.

By Fact 2.8 and (3.8), there exist $\widetilde{\varepsilon_0}, \widetilde{\gamma_0} > 0$ such that $\widetilde{\varepsilon_0} \le \varepsilon_0$ and, for all x with $x-\bar{x} \in M^{\perp}\cap\mathbb{B}(0,\widetilde{\varepsilon_0})$,

$$\max_{i\in I} g_i(x) \le 1$$

and

$$\begin{aligned} \|x-\bar{x}\| &\le \widetilde{\gamma_0}\Big(\max_{i\in I} g_i(x)\Big)^{\frac{1}{\beta(n-k-1)d^{n-k}}} \\ &= \widetilde{\gamma_0}\Big(\max_{i\in I}[g_i(x)]_+\Big)^{\frac{1}{\beta(n-k-1)d^{n-k}}}. \end{aligned} \tag{3.11}$$

[Distance estimation on M] Set $r := \begin{cases} \max\{\frac{\sum_{j\in I\setminus\{i\}}\alpha_j}{\alpha_i} \mid i\in I\} > 0, & \text{if } |I|\geq 2; \\ 1, & \text{otherwise.} \end{cases}$.
Thus $r \geq 1$. Note that $\sum_{i\in I}\alpha_i g_i(x) \geq 0$, we have for each $i \in I$

$$\max_{i\in I}[g_i(x)]_+ \geq g_i(x) \geq -\frac{\sum_{j\in I\setminus\{i\}}\alpha_j g_j(x)}{\alpha_i} \geq -r\max_{i\in I}[g_i(x)]_+ \ .$$

Hence we have $|g_i(x)| \leq r\max_{i\in I}[g_i(x)]_+$. This together with (3.9) implies that

$$\|x-\bar{x}\| \leq \gamma_0 r^2|I|\Big(\max_{i\in I}[g_i(x)]_+\Big)^{\frac{2}{\kappa(n-k,2d)}}, \quad \forall x-\bar{x}\in M^\perp\cap\mathbb{B}(0,\varepsilon_0).$$

Combining this with (3.11), we see that, for every $x-\bar{x}\in M^\perp\cap\mathbb{B}(0,\widetilde{\varepsilon}_0)$,

$$\|x-\bar{x}\| \leq \big(\gamma_0 r^2|I|+\widetilde{\gamma}_0\big)\Big(\max_{i\in I}[g_i(x)]_+\Big)^{\max\{\frac{2}{\kappa(n-k,2d)},\frac{1}{\beta(n-k-1)d^{n-k}}\}}. \tag{3.12}$$

We now consider two cases.

Case 1: $\dim M = \{0\}$.

We have $D = S = \{\bar{x}\}$. Thus $M = 0$ and $M^\perp = \mathbb{R}^n$. We can assume that $\max_{i\in I}[g_i(x)]_+ \leq 1$ for all $x-\bar{x}\in\mathbb{B}(0,\widetilde{\varepsilon}_0)$. Then by (3.12), we have

$$\mathrm{dist}(x,S) = \|x-\bar{x}\| \leq \big(\gamma_0 r^2|I|+\widetilde{\gamma}_0\big)\Big(\max_{i\in I}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}}, \quad \forall\|x-\overline{x}\|\leq\widetilde{\varepsilon}_0.$$

Case 2: $k = \dim M \geq 1$.

Since $\dim M = k$, there exists a full rank matrix $Q\in\mathbb{R}^{n\times k}$ such that $Q(\mathbb{R}^k) = M$. For each $u\in M$ and $i\in I$, (3.6) implies that

$$g_i^\infty(u) = \lim_{t\to\infty}\frac{g_i(\bar{x}+tu)-g_i(\bar{x})}{t} = 0.$$

Then, Fact 2.14 implies that

$$g_i(x+u) = g_i(x) \quad \text{for all } x\in\mathbb{R}^n, u\in M, i\in I. \tag{3.13}$$

Since $S\subseteq D = \bar{x}+M$, it follows that

$$S = \{x\in\bar{x}+M\in\mathbb{R}^n \mid g_i(x)\leq 0, i\notin I\} = \bar{x}+Q(\widehat{S}),$$

where $\widehat{S} := \{a\in\mathbb{R}^k \mid g_i(\bar{x}+Qa)\leq 0, i\notin I\}$.

Note that $0\in\widehat{S}$. The induction hypothesis implies that there exist $\widetilde{\varepsilon}_1,\widetilde{\gamma}_1>0$ such that $\max_{i\notin I}[g_i(\bar{x}+Qa)]_+\leq 1$ and

$$\mathrm{dist}(a,\widehat{S}) \leq \widetilde{\gamma}_1\Big(\max_{i\notin I}[g_i(\bar{x}+Qa)]_+\Big)^{\max\left\{\frac{2}{\kappa(k,2d)},\frac{1}{\beta(k-1)d^k}\right\}} \quad \text{for all } \|a\|\leq\widetilde{\varepsilon}_1.$$

This implies that there exist $\epsilon_2, \gamma_2 > 0$ such that

$$\text{(3.14)} \quad \operatorname{dist}(x,S) \le \gamma_2\Big(\max_{i\notin I}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(k,2d)},\frac{1}{\beta(k-1)d^k}\right\}} \quad \text{for all } x-\bar{x}\in M\cap\mathbb{B}(0,\varepsilon_2).$$

[Combining the estimation and simplification] Now let $\varepsilon \le \min\{\widetilde{\varepsilon_0},\varepsilon_2\}$ be such that $\max_{1\le i\le m}[g_i(x)]_+ \le 1$ for all $x\in\mathbb{B}(\bar{x},\varepsilon)$. Let K be a compact set contains $\mathbb{B}(\bar{x},\varepsilon)\cup\mathbb{B}(0,\varepsilon)$. Denote the Lipschitz constant of g_i over K by L_i, i.e., $|g_i(x_1)-g_i(x_2)|\le L_i\|x_1-x_2\|$ for all $x_1,x_2\in K$. Set $L:=\max_{1\le i\le m}L_i$ and $\gamma:=\max\{\gamma_0 r^2|I|+\widetilde{\gamma_0},\gamma_2\}$.

To see the conclusion, we only need to show that for any $x\in\mathbb{B}(\bar{x},\varepsilon)$,

$$\operatorname{dist}(x,S)\le c\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}},$$

where $c:=2\gamma+L\gamma^2$. To see this, let us fix an arbitrary $x\in\mathbb{B}(\bar{x},\varepsilon)$. Note that $\mathbb{R}^n = M+M^{\perp}$. Then, one can decompose $x-\bar{x}=u+v$ for some $u\in M\cap\mathbb{B}(0,\varepsilon)$ and $v\in M^{\perp}\cap\mathbb{B}(0,\varepsilon)$. This together with (3.14) and (3.12) implies that

$$\text{(3.15)} \qquad \operatorname{dist}(u+\bar{x},S)\le\gamma\Big(\max_{i\notin I}[g_i(u+\bar{x})]_+\Big)^{\max\left\{\frac{2}{\kappa(k,2d)},\frac{1}{\beta(k-1)d^k}\right\}}$$

$$\text{(3.16)} \qquad \|v\|\le\gamma\Big(\max_{i\in I}[g_i(v+\bar{x})]_+\Big)^{\max\left\{\frac{2}{\kappa(n-k,2d)},\frac{1}{\beta(n-k-1)d^{n-k}}\right\}}.$$

Therefore,

$$\begin{aligned}
&\operatorname{dist}(x,S)\le\operatorname{dist}(u+\bar{x},S)+\|x-(u+\bar{x})\|\\
&=\operatorname{dist}(u+\bar{x},S)+\|v\|\\
&\le\gamma\Big(\max_{i\notin I}[g_i(u+\bar{x})]_+\Big)^{\max\left\{\frac{2}{\kappa(k,2d)},\frac{1}{\beta(k-1)d^k}\right\}}+\gamma\Big(\max_{i\in I}[g_i(v+\bar{x})]_+\Big)^{\max\left\{\frac{2}{\kappa(n-k,2d)},\frac{1}{\beta(n-k-1)d^{n-k}}\right\}}\\
&=\gamma\Big(\max_{i\notin I}[g_i(u+\bar{x})]_+\Big)^{\max\left\{\frac{2}{\kappa(k,2d)},\frac{1}{\beta(k-1)d^k}\right\}}+\gamma\Big(\max_{i\in I}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}}\\
&\le\gamma\Big(\max_{1\le i\le m}[g_i(u+\bar{x})]_+\Big)^{\max\left\{\frac{2}{\kappa(k,2d)},\frac{1}{\beta(k-1)d^k}\right\}}+\gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}},
\end{aligned}$$

where the second equality follows by (3.13) and $v+\bar{x}+u=x$. Note that

$$\begin{aligned}
&\Big|\max_{1\le i\le m}[g_i(u+\bar{x})]_+-\max_{1\le i\le m}[g_i(x)]_+\Big|\le\max_{1\le i\le m}|g_i(u+\bar{x})-g_i(x)|\\
&\le L\|u+\bar{x}-x\|=L\|v\|\\
&\le L\gamma\Big(\max_{i\in I}[g_i(v+\bar{x})]_+\Big)^{\max\left\{\frac{2}{\kappa(n-k,2d)},\frac{1}{\beta(n-k-1)d^{n-k}}\right\}} \quad \text{(by (3.16))}
\end{aligned}$$

$$\text{(3.17)} \quad = L\gamma\Big(\max_{i\in I}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n-k,2d)},\frac{1}{\beta(n-k-1)d^{n-k}}\right\}} \quad \text{(by 3.13) and } v+\bar{x}+u=x).$$

As $\max_{1\le i\le m}[g_i(x)]_+ \le 1$ for all $x \in \mathbb{B}(\bar{x},\varepsilon)$, it follows that

$$
\begin{aligned}
\operatorname{dist}(x,S) &\le \gamma\Big(\max_{1\le i\le m}[g_i(u+\bar{x})]_+\Big)^{\frac{2}{\kappa(k,2d)}} \quad + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}} \\
&\le \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+ + L\gamma(\max_{i\in I}[g_i(x)]_+^{\frac{2}{\kappa(n-k,2d)}})\Big)^{\frac{2}{\kappa(k,2d)}} \quad \text{(by (3.17))} \\
&\quad + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}} \\
&\le \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+^{\frac{2}{\kappa(n-k,2d)}} + L\gamma(\max_{i\in I}[g_i(x)]_+^{\frac{2}{\kappa(n-k,2d)}})\Big)^{\frac{2}{\kappa(k,2d)}} \\
&\quad + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}} \\
&\le (\gamma+L\gamma^2)\max_{1\le i\le m}[g_i(x)]_+^{\frac{4}{\kappa(n-k,2d)\cdot\kappa(k,2d)}} + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}} \\
&= (\gamma+L\gamma^2)\max_{1\le i\le m}[g_i(x)]_+^{\frac{4}{\left((2d-1)^{n-k}+1\right)\cdot\left((2d-1)^k+1\right)}} + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}} \\
&\le (\gamma+L\gamma^2)\max_{1\le i\le m}[g_i(x)]_+^{\frac{2}{(2d-1)^n+1}} + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}}.
\end{aligned} \tag{3.18}
$$

Similarly, we also have

$$
\begin{aligned}
\operatorname{dist}(x,S) &\le \gamma\Big(\max_{1\le i\le m}[g_i(u+\bar{x})]_+\Big)^{\frac{1}{\beta(k-1)d^k}} + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}} \\
&\le \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+ + L\gamma(\max_{i\in I}[g_i(x)]_+^{\frac{1}{\beta(n-k-1)d^{n-k}}})\Big)^{\frac{1}{\beta(k-1)d^k}} \quad \text{(by (3.17))} \\
&\quad + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}} \\
&\le \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+^{\frac{1}{\beta(n-k-1)d^{n-k}}} + L\gamma(\max_{i\in I}[g_i(x)]_+^{\frac{1}{\beta(n-k-1)d^{n-k}}})\Big)^{\frac{1}{\beta(k-1)d^k}} \\
&\quad + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}} \\
&= (\gamma+L\gamma^2)\max_{1\le i\le m}[g_i(x)]_+^{\frac{1}{\left(\beta(n-k-1)d^{n-k}\right)\cdot\left(\beta(k-1)d^k\right)}} + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}} \\
&\le (\gamma+L\gamma^2)\max_{1\le i\le m}[g_i(x)]_+^{\frac{1}{\beta(n-1)d^n}} + \gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}},
\end{aligned}
$$

where the last inequality was obtained by the Chu-Vandermonde identity.

In combination with (3.18) we obtain

$$\begin{aligned}\mathrm{dist}(x,S) &\le (\gamma+L\gamma^2)\max_{1\le i\le m}[g_i(x)]_+^{\max\left\{\frac{2}{(2d-1)^n+1},\frac{1}{\beta(n-1)d^n}\right\}} \quad +\gamma\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\}}\\ &= (2\gamma+L\gamma^2)\max_{1\le i\le m}[g_i(x)]_+^{\max\left\{\frac{2}{(2d-1)^n+1},\frac{1}{\beta(n-1)d^n}\right\}}.\end{aligned}$$

This completes the proof. $\square$

As a corollary, we obtain a local error bound result which is independent of the partition of the index set.

Corollary 3.4 *Let g_i be convex polynomials on $\mathbb{R}^n$ with degree at most d for every $i = 1,\cdots,m$. Let $S := \{x \in \mathbb{R}^n \mid g_i(x) \le 0, i = 1,\cdots,m\}$ and $\bar{x} \in S$. Then there exist $c,\varepsilon > 0$ such that*

$$\mathrm{dist}(x,S) \le c\Big(\max_{1\le i\le m}[g_i(x)]_+\Big)^{\tau} \quad \textit{whenever} \quad \|x-\overline{x}\| \le \varepsilon,$$

where $[a]_+ := \max\{a,0\}$, $\tau := \max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\} = \frac{1}{\min\left\{\frac{(2d-1)^n+1}{2},\beta(n-1)d^n\right\}}$ and $\beta(n-1)$ is the central binomial coefficient with respect to $n-1$.

Proof. Choose ε small enough so that $\max_{1\le i\le m}[g_i(x)]_+ \le 1$. Then, the conclusion follows immediately from the preceding Theorem 3.3 by noting that $[g_i(x)]_+ \le \big(\max_{1\le i\le m}[g_i(x)]_+\big)^{\frac{2}{\kappa(n,2d)}}$ for each $i = 1,\ldots,m$. $\square$

Remark 3.5 (Discussion on the exponent) Let g_i be convex polynomials on $\mathbb{R}^n$ with degree at most d for every $i = 1,\cdots,m$. Let $S := \{x \in \mathbb{R}^n \mid g_i(x) \le 0, i = 1,\cdots,m\}$ and $\bar{x} \in S$. We now form some discussion on the exponent in our local error bound results.

(1) Theorem 3.3 shows that in the case when $d = 1$ or there exists $x_0 \in \mathbb{R}^n$ such that $g_i(x_0) < 0$, $i = 1,\ldots,m$, we indeed obtain the Lipschitz type local error bound. That is to say, in these cases, there exist $c,\varepsilon > 0$ such that

$$\mathrm{dist}(x,S) \le c\max_{1\le i\le m}[g_i(x)]_+ \quad \textit{whenever} \quad \|x-\overline{x}\| \le \varepsilon,$$

where $[a]_+ := \max\{a,0\}$. To see this, if $d = 1$, then $\frac{2}{\kappa(n,2d)} = 1$ and $\frac{1}{\beta(n-1)d^n} \ge 1$. So, $\tau = \max\left\{\frac{2}{\kappa(n,2d)},\frac{1}{\beta(n-1)d^n}\right\} = 1$ and the conclusion follows immediately from Theorem 3.3. On the other hand, if there exists $x_0 \in \mathbb{R}^n$ such that $g_i(x_0) < 0$, $i = 1,\ldots,m$, then, $J_1 = \emptyset$. So, the conclusion follows immediately from the same theorem, and also recovers Corollary 2.11.

(2) In the case when $n = 1$, we see that $\frac{2}{\kappa(n,2d)} = \frac{1}{\beta(n-1)d^n} = \frac{1}{d}$. So, when each g_i is a univariate convex polynomial, then there exist $c,\varepsilon > 0$ such that

$$\mathrm{dist}(x,S) \le c\big(\max_{1\le i\le m}[g_i(x)]_+\big)^{\frac{1}{d}} \quad \textit{whenever} \quad \|x-\overline{x}\| \le \varepsilon,$$

which seems to match what one might expect in this case.

(3) On the other hand, in general, our estimation on the exponent will not be optimal.

For example, if the inequality system consists of one single convex polynomial, Fact 2.9 shows that the exponent can be set as $\frac{1}{(d-1)^n+1}$ while our results produce a weaker exponent $\max\left\{\frac{2}{(2d-1)^n+1}, \frac{1}{\beta(n-1)d^n}\right\}$. An interesting feature of the exponent $\frac{1}{(d-1)^n+1}$ in Fact 2.9 is that, in the convex quadratic case, it collapses to $\frac{1}{2}$ which is independent of the dimension of the underlying space and agrees with the known result presented in [35].

By contrast, our estimate $\max\left\{\frac{2}{3^n+1}, \frac{1}{\beta(n-1)2^n}\right\}$ which depends heavily on the dimension n. Moreover, as indicated in Example 3.1, the best possible exponent might be $\frac{1}{d^n}$ (see [29] for some relevant discussion regarding the best possible exponent for general nonconvex polynomial system). It would be interesting to know how one could improve our estimate here.

Making better sense of these estimates will be one of our future research topics. $\Diamond$

Given $D \subseteq \mathbb{R}^n$, we set $\mathrm{dist}^r(\cdot, D) := \big(\mathrm{dist}(\cdot, D)\big)^r$ for every $r \in \mathbb{R}$.

Theorem 3.6 (Hölderian regularity) *Let $\theta > 0$ and $K \subseteq \mathbb{R}^n$ be a compact set. Then there exists $c > 0$ such that*

$$\mathrm{dist}^\theta(x, C) \le c\left(\sum_{i=1}^m \mathrm{dist}^\theta(x, C_i)\right)^\tau, \quad \forall x \in K,$$

where $\tau := \frac{1}{\min\left\{\frac{(2d-1)^n+1}{2}, \beta(n-1)d^n\right\}}$ and $\beta(n-1)$ is the central binomial coefficient with respect to $n-1$ which is given by $\binom{n-1}{[(n-1)/2]}$.

Proof. To see the conclusion, we only need to show that for each $\bar{x} \in \mathbb{R}^n$, there exist $c, \varepsilon > 0$ such that

$$\mathrm{dist}^\theta(x, C) \le c\left(\sum_{i=1}^m \mathrm{dist}^\theta(x, C_i)\right)^\tau, \quad \text{for all } \|x - \bar{x}\| \le \varepsilon. \tag{3.19}$$

Indeed, granting this and fixing a compact set K, then for any $\bar{x} \in K$ there exist $c_{\bar{x}}, \varepsilon_{\bar{x}} > 0$ such that

$$\mathrm{dist}^\theta(x, C) \le c_x\left(\sum_{i=1}^m \mathrm{dist}^\theta(x, C_i)\right)^\tau, \quad \text{for all } \|x - \bar{x}\| \le \varepsilon_x.$$

As K is compact and $\bigcup_{\bar{x}\in K} \mathbb{B}(\bar{x}; \varepsilon_{\bar{x}}) \supseteq K$, we can find finitely many points $\bar{x}_1, \cdots, \bar{x}_s \in K$, $s \in \mathbb{N}$, such that $\bigcup_{i=1}^s \mathbb{B}(\bar{x}_i; \varepsilon_{\bar{x}_i}) \supseteq K$. Let $c := \max\{c_{\bar{x}_1}, \cdots, c_{\bar{x}_s}\}$. Then, for any $x \in K$, there exists $i_0 \in \{1, \cdots, s\}$ such that $x \in \mathbb{B}(\bar{x}_{i_0}; \varepsilon_{\bar{x}_{i_0}})$, and hence

$$\mathrm{dist}^\theta(x, C) \le c_{\bar{x}_{i_0}}\left(\sum_{i=1}^m \mathrm{dist}^\theta(x, C_i)\right)^\tau \le c\left(\sum_{i=1}^m \mathrm{dist}^\theta(x, C_i)\right)^\tau.$$

We now show (3.19) holds. Fix $\bar{x} \in \mathbb{R}^n$. We consider two cases.

Case 1: $\bar{x} \notin C$.

Then there exist $\varepsilon_1, \eta, M > 0$ such that

$$\sum_{i=1}^{m} \mathrm{dist}^{\theta}(x, C_i) \geq \eta \text{ and } \mathrm{dist}^{\theta}(x, C) \leq M \quad \text{for all } \|x - \bar{x}\| \leq \varepsilon_1.$$

Therefore, $\mathrm{dist}^{\theta}(x, C) \leq M = \frac{M}{\eta^{\tau}} \eta^{\tau} \leq \frac{M}{\eta^{\tau}} \Big(\sum_{i=1}^{m} \mathrm{dist}^{\theta}(x, C_i) \Big)^{\tau}$ for all $\|x - \bar{x}\| \leq \varepsilon_1$ and hence, (3.19) holds.

Case 2: $\bar{x} \in C$.

We have

$$C = \Big\{ x \in \mathbb{R}^n \mid g_{i,1}(x) \leq 0, g_{i,2}(x) \leq 0, \cdots, g_{i,\gamma_i}(x) \leq 0, \quad i = 1, \cdots, m \Big\}.$$

By Corollary 3.4, there exist positive constants c_0 and δ such that

$$\mathrm{dist}(x, C) \leq c_0^{\frac{1}{\theta}} \bigg(\max_{1 \leq i \leq m} \big\{ [g_{i,1}(x)]_+, \cdots, [g_{i,\gamma_i}(x)]_+ \big\} \bigg)^{\tau}, \quad \forall \|x - \bar{x}\| \leq \delta.$$

Hence

$$\mathrm{dist}^{\theta}(x, C) \leq c_0 \bigg(\max_{1 \leq i \leq m} \big\{ [g_{i,1}(x)]_+, \cdots, [g_{i,\gamma_i}(x)]_+ \big\} \bigg)^{\theta\tau}, \quad \forall \|x - \bar{x}\| \leq \delta. \tag{3.20}$$

Now we claim that there exists $\beta > 0$ such that

$$\bigg(\max_{1 \leq i \leq m} \big\{ g_{i,1}(x)]_+, \cdots, [g_{i,\gamma_i}(x)]_+ \big\} \bigg)^{\theta} \leq \beta \sum_{i=1}^{m} \mathrm{dist}^{\theta}(x, C_i), \quad \forall \|x - \bar{x}\| \leq \delta. \tag{3.21}$$

Suppose to the contrary that there exists a sequence $(x_k)_{k \in \mathbb{N}}$ in $\mathbb{B}(\bar{x}, \delta)$ such that

$$\bigg(\max_{1 \leq i \leq m} \big\{ [g_{i,1}(x_k)]_+, \cdots, [g_{i,\gamma_i}(x_k)]_+ \big\} \bigg)^{\theta} > k \sum_{i=1}^{m} \mathrm{dist}^{\theta}(x_k, C_i), \quad \forall k \in \mathbb{N}. \tag{3.22}$$

Without loss of generality, we can assume that $g_{i,1}, g_{i,2}, \cdots g_{i,\gamma_i}$ have the Lipschitz constant $L > 0$ on $\mathbb{B}(\bar{x}, \delta)$ for every $i = 1, 2, \cdots, m$. Then, there exists a subsequence $(x_{k_l})_{l \in \mathbb{N}}$ of $(x_k)_{k \in \mathbb{N}}$, $1 \leq i_0 \leq m$ and $1 \leq j_0 \leq \gamma_{i_0}$ such that

$$\max_{1 \leq i \leq m} \big\{ [g_{i,1}(x_{k_l})]_+, \cdots, [g_{i,\gamma_i}(x_{k_l})]_+ \big\} = [g_{i_0,j_0}(x_{k_l})]_+, \quad \forall l \in \mathbb{N}.$$

It follows from (3.22) that

$$\big([g_{i_0,j_0}(x_{k_l})]_+ \big)^{\theta} > k_l \sum_{i=1}^{m} \mathrm{dist}^{\theta}(x_{k_l}, C_i), \quad \forall l \in \mathbb{N}. \tag{3.23}$$

Then $[g_{i_0,j_0}(x_{k_l})]_+ = g_{i_0,j_0}(x_{k_l})$ and hence for every $l \in \mathbb{N}$,

(3.24)

$$\big(g_{i_0,j_0}(x_{k_l})\big)^\theta > k_l \sum_{i=1}^m \mathrm{dist}^\theta(x_{k_l}, C_i) = k_l \sum_{i=1}^m \big\|x_{k_l} - P_i(x_{k_l})\big\|^\theta \ge k_l \big\|x_{k_l} - P_{i_0}(x_{k_l})\big\|^\theta.$$

Since $P_{i_0}(x_{k_l}) \in C_{i_0}$, we have $g_{i_0,j_0}\big(P_{i_0}(x_{k_l})\big) \le 0$ and $\|x_{k_l} - P_{i_0}(x_{k_l})\| \le \|x_{k_l} - \bar{x}\| < \delta$ by $\bar{x} \in C_{i_0}$. Combining this with (3.24), we have

$$L^\theta \big\|x_{k_l} - P_{i_0}(x_{k_l})\big\|^\theta \ge \Big(g_{i_0,j_0}(x_{k_l}) - g_{i_0,j_0}\big(P_{i_0}(x_{k_l})\big)\Big)^\theta > k_l \big\|x_{k_l} - P_{i_0}(x_{k_l})\big\|^\theta, \quad \forall l \in \mathbb{N}.$$

Hence we have $L^\theta > k_l$ for every $l \in \mathbb{N}$, this contradicts the fact that $k_l \longrightarrow +\infty$. Thus, (3.21) holds.

Combining (3.21) and (3.20), we see that

$$\mathrm{dist}^\theta(x, C) \le c_0 \beta^\tau \Big(\sum_{i=1}^m \mathrm{dist}^\theta(x, C_i)\Big)^\tau, \quad \forall \|x - \bar{x}\| \le \delta,$$

and so the conclusion follows. □

4 Convergence rate for the cyclic projection algorithm

In this section, we derive explicit convergence rate of the cyclic projection algorithm applied to finite intersections of semi-algebraic convex sets.

Before we come to our main result, we need the following useful lemma.

Lemma 4.1 (Recurrence relationships) *Let $p > 0$, and let $\{\delta_k\}_{k=0}^\infty$ and $\{\beta_k\}_{k=0}^\infty$ be two sequences of nonnegative numbers satisfying the conditions*

$$\beta_{k+1} \le \beta_k (1 - \delta_k \beta_k^p) \ \ as \ \ k = 0, 1, \ldots.$$

Then there is a number $\gamma > 0$ such that

(4.1)
$$\beta_k \le \left(\beta_0^{-p} + p \sum_{i=0}^{k-1} \delta_i\right)^{-\frac{1}{p}} \ \ \textit{for all} \ \ k \in \mathbb{N}.$$

We use the convention that $\frac{1}{0} = +\infty$. In particular, we have $\lim_{k\to\infty} \beta_k = 0$ whenever $\sum_{k=0}^\infty \delta_k = \infty$.

Proof. It follows from our assumption that

$$0 \le \beta_{i+1} \le \beta_i \le \cdots \le \beta_0 \ \ \text{and} \ \ \delta_i \beta_i^{p+1} \le \beta_i - \beta_{i+1} \ \ \text{as} \ \ i \in \mathbb{N}.$$

Fix $k \in \mathbb{N}$. We consider two cases.

Case 1: $\beta_k = 0$.

Clearly, (4.1) holds.

Case 1: $\beta_k \neq 0$.

Thus $\beta_k > 0$ and hence $\beta_i > 0$ for every $i \leq k$. Define the nonincreasing function $h : \mathbb{R}_{++} \to \,]-\infty, +\infty]$ by $h(x) := x^{-(p+1)}$. As $\delta_i\, h(\beta_i)^{-1} = \delta_i \beta_i^{p+1} \leq \beta_i - \beta_{i+1}$, then we get

$$\delta_i \leq (\beta_i - \beta_{i+1}) h(\beta_i) \leq \int_{\beta_{i+1}}^{\beta_i} h(x)dx = \frac{\beta_{i+1}^{-p} - \beta_i^{-p}}{p}.$$

This implies that

$$\beta_{i+1}^{-p} - \beta_i^{-p} \geq p\delta_i \ \text{ for all } \ i \in \mathbb{N} \cup \{0\}. \tag{4.2}$$

Now fix any $k \in \mathbb{N}$ and, summing (4.2) from $i = 0$ to $i = k-1$, we get

$$\beta_k^{-p} - \beta_0^{-p} \geq p \sum_{i=0}^{k-1} \delta_i.$$

which implies the conclusion in (4.1). □

We are now ready for our main result: Theorem 4.2. The proof follows in part that of [26, Lemmas 3&4], one may also consult [8].

Theorem 4.2 (Cyclic convergence rate) *Suppose that $d > 1$. Let $x_0 \in \mathbb{R}^n$ and the sequence of cyclic projections, $(x_k)_{k\in\mathbb{N}}$, be defined by*

$$x_1 := P_1 x_0,\, x_2 := P_2 x_1,\, \cdots,\, x_m := P_m x_{m-1},\, x_{m+1} := P_1 x_m \ldots.$$

Then x_k converges to $x_\infty \in C$, and there exists $M > 0$ such that

$$\|x_k - x_\infty\| \leq M \frac{1}{k^\rho}, \quad \forall k \in \mathbb{N},$$

where $\rho := \frac{1}{\min\left\{(2d-1)^n - 1,\, 2\beta(n-1)d^n - 2\right\}}$ and $\beta(n-1)$ is the central binomial coefficient with respect to $n-1$ which is given by $\binom{n-1}{[(n-1)/2]}$.

Proof. We denoted by $\alpha_i := (i \mod m) + 1, \forall i \in \mathbb{N}$. Thus $x_{k+1} = P_{C_{\alpha_k}} x_k$. By Fact 2.18, there exists $x_\infty \in C$ such that $x_k \longrightarrow x_\infty$.

We first follow closely the proofs of [26, Lemmas 3&4] to get that

$$\operatorname{dist}^2(x_k, C) - \operatorname{dist}^2(x_{k+1}, C) \geq \operatorname{dist}^2(x_k, C_{\alpha_k}), \quad \forall k \in \mathbb{N}. \tag{4.3}$$

We have

$$
\begin{aligned}
&\operatorname{dist}^2(x_k, C) - \operatorname{dist}^2(x_{k+1}, C) \geq \|x_k - P_C x_k\|^2 - \|x_{k+1} - P_C x_k\|^2 \\
&= \|x_k - P_C x_k\|^2 - \|P_{\alpha_k} x_k - P_C x_k\|^2 = \|x_k - P_C x_k\|^2 - \big\|P_{\alpha_k} x_k - x_k + x_k - P_C x_k\big\|^2 \\
&= \|x_k - P_C x_k\|^2 - \|P_{\alpha_k} x_k - x_k\|^2 - \|x_k - P_C x_k\|^2 + 2\big\langle x_k - P_{\alpha_k} x_k, x_k - P_C x_k\big\rangle \\
&\geq -\|x_k - P_{\alpha_k} x_k\|^2 + 2\big\langle x_k - P_{\alpha_k} x_k, x_k - P_{\alpha_k} x_k + P_{\alpha_k} x_k - P_C x_k\big\rangle \\
&= -\|x_k - P_{\alpha_k} x_k\|^2 + 2\|x_k - P_{\alpha_k} x_k\|^2 + 2\big\langle x_k - P_{\alpha_k} x_k, P_{\alpha_k} x_k - P_C x_k\big\rangle \\
&\geq \|x_k - P_{\alpha_k} x_k\|^2 = \operatorname{dist}^2(x_k, C_{\alpha_k}), \quad \forall k \in \mathbb{N}.
\end{aligned}
$$

Hence (4.3) holds.

Next we claim that for every $i \in \mathbb{N}$

$$
\operatorname{dist}(x_k, C_{\alpha_i}) \leq \operatorname{dist}(x_k, C_{\alpha_k}) + \operatorname{dist}(x_{k+1}, C_{\alpha_{k+1}}) + \ldots + \operatorname{dist}(x_{k+m-1}, C_{\alpha_{k+m-1}}). \tag{4.4}
$$

To see this, note that there exists $i_0 \leq m-1$ such that $\alpha_{i_0+k} = \alpha_i$. Then, we have

$$
\begin{aligned}
&\operatorname{dist}(x_k, C_{\alpha_i}) = \operatorname{dist}(x_k, C_{\alpha_{i_0+k}}) \leq \|x_k - x_{i_0+k}\| + \big\|x_{i_0+k} - P_{\alpha_{i_0+k}} x_{i_0+k}\big\| \\
&\leq \|x_k - x_{k+1}\| + \cdots + \|x_{i_0+k-1} - x_{i_0+k}\| + \big\|x_{i_0+k} - P_{\alpha_{i_0+k}} x_{i_0+k}\big\| \\
&= \|x_k - P_{\alpha_k} x_k\| + \cdots + \|x_{i_0+k-1} - P_{\alpha_{i_0+k-1}} x_{i_0+k-1}\| + \big\|x_{i_0+k} - P_{\alpha_{i_0+k}} x_{i_0+k}\big\| \\
&= \operatorname{dist}(x_k, C_{\alpha_k}) + \operatorname{dist}(x_{k+1}, C_{\alpha_{k+1}}) + \cdots + \operatorname{dist}(x_{i_0+k}, C_{\alpha_{i_0+k}}) \\
&\leq \operatorname{dist}(x_k, C_{\alpha_k}) + \operatorname{dist}(x_{k+1}, C_{\alpha_{k+1}}) + \cdots + \operatorname{dist}(x_{k+m-1}, C_{\alpha_{k+m-1}}) \quad (\text{by } i_0 \leq m-1).
\end{aligned}
$$

Hence (4.4) holds.

Thus by (4.4),

$$
\begin{aligned}
&\operatorname{dist}^2(x_k, C_{\alpha_i}) \leq \Big(m \max_{k \leq i \leq k+m-1} \operatorname{dist}(x_i, C_{\alpha_i})\Big)^2 \\
&\leq m^2 \Big(\operatorname{dist}^2(x_k, C_{\alpha_k}) + \operatorname{dist}^2(x_{k+1}, C_{\alpha_{k+1}}) + \ldots + \operatorname{dist}^2(x_{k+m-1}, C_{\alpha_{k+m-1}})\Big).
\end{aligned} \tag{4.5}
$$

By Theorem 3.6 and Fact 2.18, there exists $c_0 > 0$ such that

$$
\operatorname{dist}^2(x_k, C) \leq c_0 \Big(\sum_{i=1}^{m} \operatorname{dist}^2(x_k, C_i)\Big)^{\frac{1}{r}}, \quad \forall k \in \mathbb{N}.
$$

where $r := \min\big\{\frac{(2d-1)^n+1}{2}, \beta(n-1)d^n\big\}$.

Then by (4.5), for every $k \in \mathbb{N}$,

$$\begin{aligned}
&\frac{1}{c_0^r}\operatorname{dist}^{2r}(x_k, C) \\
&\le \sum_{i=1}^{m}\operatorname{dist}^2(x_k, C_i) \le m \max_{1\le i\le m}\operatorname{dist}^2(x_k, C_i) \\
&\le m^3\Big(\operatorname{dist}^2(x_k, C_{\alpha_k}) + \operatorname{dist}^2(x_{k+1}, C_{\alpha_{k+1}}) + \ldots + \operatorname{dist}^2(x_{k+m-1}, C_{\alpha_{k+m-1}})\Big) \\
&\le m^3\sum_{i=k}^{k+m-1}\operatorname{dist}^2(x_i, C) - \operatorname{dist}^2(x_{i+1}, C) \quad \text{(by (4.3))} \\
&= m^3\Big(\operatorname{dist}^2(x_k, C) - \operatorname{dist}^2(x_{k+m}, C)\Big). \qquad (4.6)
\end{aligned}$$

Thus we have

$$\operatorname{dist}^2(x_{k+m}, C) \le \operatorname{dist}^2(x_k, C) - \frac{1}{m^3 c_0^r}\operatorname{dist}^{2r}(x_k, C). \tag{4.7}$$

Now fix $k_0 \in \mathbb{N}$. Let $\beta_i := \operatorname{dist}^2(x_{k_0+im}, C)$, $\forall i \in \mathbb{N}\cup\{0\}$. Then (4.7) shows that

$$\beta_{i+1} \le \beta_i - \frac{1}{m^3 c_0^r}\beta_i^r = \beta_i\Big(1 - \frac{1}{m^3 c_0^r}\beta_i^{r-1}\Big). \tag{4.8}$$

By Lemma 4.1,

$$\operatorname{dist}^2(x_{k_0+im}, C) = \beta_i \le \left(\beta_0^{1-r} + (r-1)i\frac{1}{m^3 c_0^r}\right)^{-\frac{1}{r-1}}, \quad \forall i \in \mathbb{N}.$$

Thus there exits $M_0 > 0$ such that

$$\operatorname{dist}(x_{k_0+im}, C) \le M_0\frac{1}{\sqrt[2(r-1)]{i}}, \quad \forall i \in \mathbb{N}.$$

Hence we have there exits $M_1 > 0$ such that

$$\operatorname{dist}(x_k, C) \le M_1\frac{1}{\sqrt[2(r-1)]{k}} = M_1\frac{1}{k^\rho}, \quad \forall k \in \mathbb{N}, \tag{4.9}$$

where $\rho := \frac{1}{\min\left\{(2d-1)^n-1,\, 2\beta(n-1)d^n-2\right\}}$.

We have

$$\|x_k - x_\infty\| \le \|x_k - P_C(x_k)\| + \|P_C(x_k) - x_\infty\| = \operatorname{dist}(x_k, C) + \|P_C(x_k) - x_\infty\|. \tag{4.10}$$

By [26, Lemma 3],

$$\|x_{k+l} - P_C(x_k)\| \le \|x_k - P_C(x_k)\| = \operatorname{dist}(x_k, C), \quad \forall l \in \mathbb{N}.$$

Taking $l \longrightarrow \infty$ in the above inequality, we obtain that

$$\|x_\infty - P_C(x_k)\| \le \operatorname{dist}(x_k, C). \tag{4.11}$$

Combining (4.10), (4.11) and (4.9),

$$\|x_k - x_\infty\| \le 2\,\mathrm{dist}(x_k, C) \le 2M_1 \frac{1}{k^\rho}, \quad \forall k \in \mathbb{N}. \tag{4.12}$$

Thus, the conclusion follows by letting $M := 2M_1$. $\square$

As a corollary, in the case of two sets with nonempty intersection, we obtain the following estimate on the convergence rate of the alternating projection algorithm.

Corollary 4.3 (Alternating convergence rate) *Let A, B be defined as in Theorem 4.7. Suppose that $A \cap B \neq \emptyset$ and $d > 1$. Let the sequence $\{(a_k, b_k)\}$ be generated by the alternating projection algorithm (see Subsection 4.1). Then, $a_k, b_k \longrightarrow c \in A \cap B$. Moreover, there exists a constant $M > 0$ such that*

$$\|a_k - c\| \le M \frac{1}{k^\rho} \quad \text{and} \quad \|b_k - c\| \le M \frac{1}{k^\rho}.$$

where $\rho := \frac{1}{\min\left\{(2d-1)^n - 1,\, 2\beta(n-1)d^n - 2\right\}}$ and $\beta(n-1)$ is the central binomial coefficient with respect to $n-1$ which is given by $\binom{n-1}{[(n-1)/2]}$.

4.1 The case of two sets with empty intersection

When $m = 2$, we can consider the general case where the intersection of these two sets is (possibly) empty.

We assume throughout this subsection that

$$\boxed{\begin{array}{l} g_i, h_j \text{ are convex polynomials with degree at most } d, \forall i = 1, 2, \cdots, m,\ j = 1, 2, \cdots, l \\ A := \{x \in \mathbb{R}^n \mid g_i(x) \le 0, i = 1, \cdots, m\} \\ B := \{x \in \mathbb{R}^n \mid h_j(x) \le 0, j = 1, \cdots, l\} \\ b_0 \in \mathbb{R}^n, \quad a_{k+1} := P_A b_k, \quad b_{k+1} := P_B a_{k+1}. \end{array}}$$

We first need the following lemma.

Lemma 4.4 *The difference $B - A$ of these two semi-algebraic sets A, B is closed.*

Proof. Let $b_k \in B$ and $a_k \in A$ be such that $b_k - a_k \longrightarrow c$. We now show that $c \in B - A$. Consider the following convex polynomial optimization problem

$$\begin{array}{lll} (P) & \min_{x \in \mathbb{R}^n, y \in \mathbb{R}^n} & \|(y - x) - c\|^2 \\ & \text{s.t.} & g_i(x) \le 0, i = 1, \cdots, m, \\ & & h_j(y) \le 0, j = 1, \cdots, l. \end{array}$$

Note that (a_k, b_k) are feasible for (P). Hence we see that $\inf(P) = 0$. By Fact 2.13, the optimal solution of (P) exists. Thus there exists $x \in A$ and $y \in B$ such that $c = y - x \in B - A$. Hence the conclusion follows. $\square$

Remark 4.5 With A and B defined as above, Fact 2.13 implies that $B - A$ is closed convex. Hence $P_{B-A}0 \neq \varnothing$. Let $v := P_{B-A}0$. Then, there exist $a \in A$ and $b \in B$ such that $v = b - a$ and hence $\operatorname{dist}(A, B) = \|v\|$. $\lozenge$

Remark 4.6 In general, the distance between two convex and semi-algebraic sets need not be attained. For instance, consider $D := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1x_2 \geq 1, x_1 \geq 0, x_2 \geq 0\}$ and $E := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = 0\}$. It is clear that D, E are convex sets while D is not a semi-algebraic convex set (as explained in Example 2.2). Clearly, $\operatorname{dist}(D, E) = 0$ but $D \cap E = \emptyset$. Thus, the distance is not attained in this case. $\lozenge$

The proof of Theorem 4.7 partially follows that of [6, Theorem 3.12].

Theorem 4.7 (Convergence rate in the infeasible case) *Assume $d > 1$. Then $a_k \longrightarrow \widetilde{a} \in A$ and $b_k \longrightarrow \widetilde{b} \in B$ with $\widetilde{b} - \widetilde{a} = v$ where $v := P_{B-A}0$. Moreover, there exists $M > 0$ such that for every $k \in \mathbb{N}$*

$$\|a_k - \widetilde{a}\| \leq M\frac{1}{k^\rho} \quad \textit{and} \quad \|b_k, -b\| \leq M\frac{1}{k^\rho}, \quad \forall k \in \mathbb{N}, \tag{4.13}$$

where $\rho := \frac{1}{\min\left\{(2d-1)^n - 1,\, 2\beta(n-1)d^n - 2\right\}}$ and $\beta(n-1)$ is the central binomial coefficient with respect to $n - 1$ which is given by $\binom{n-1}{[(n-1)/2]}$.

Proof. Lemma 4.4 implies that $A - B$ is closed. Then by Fact 2.15(i), there exist $\widetilde{a} \in A, \widetilde{b} \in B$ such that $a_k \longrightarrow \widetilde{a} \in A$ and $b_k \longrightarrow \widetilde{b} \in B$ with $\widetilde{b} - \widetilde{a} = v$. By Theorem 3.6, there exists $c_0 > 0$ such that

$$\operatorname{dist}(a_k, A \cap (B - v)) \leq c_0\big(\operatorname{dist}(a_k, A) + \operatorname{dist}(a_k, B - v)\big)^{\frac{1}{r}} = c_0 \operatorname{dist}^{\frac{1}{r}}(a_k, B - v), \tag{4.14}$$

where $r := \min\left\{\frac{(2d-1)^n + 1}{2}, \beta(n-1)d^n\right\}$. Fix $x \in A \cap (B - v)$. Note that $v = P_{B-A}0$ we have $P_B(x) = x + v$ by Fact 2.15(ii). This implies that

$$\begin{aligned}
\operatorname{dist}^2(a_k, B - v) &\leq \|a_k - (b_k - v)\|^2 = \|(a_k - x) - \big(b_k - (v + x)\big)\|^2 \\
&= \|(a_k - x) - (P_B a_k - P_B x)\|^2 \\
&\leq \|a_k - x\|^2 - \|P_B a_k - P_B x\|^2 \quad \text{(by [10, Proposition 4.8])} \\
&= \|a_k - x\|^2 - \|b_k - (x + v)\|^2 \\
&\leq \|a_k - x\|^2 - \|P_A b_k - P_A(x + v)\|^2 \\
&= \|a_k - x\|^2 - \|a_{k+1} - x\|^2 \quad \text{(by Fact 2.15(ii))}.
\end{aligned}$$

In particular, choose $x = P_{A\cap(B-v)}a_k$. Then, we have

$$\begin{aligned}\operatorname{dist}^2(a_k, B-v) &\leq \operatorname{dist}^2(a_k, A\cap(B-v)) - \|a_{k+1} - P_{A\cap(B-v)}a_k\|^2\\ &\leq \operatorname{dist}^2(a_k, A\cap(B-v)) - \operatorname{dist}^2(a_{k+1}, A\cap(B-v)).\end{aligned}$$

Combining with (4.14), we have

$$\begin{aligned}\frac{1}{c_0^{2r}}\operatorname{dist}^{2r}(a_k, A\cap(B-v)) &\leq \operatorname{dist}^2(a_k, (B-v))\\ &\leq \operatorname{dist}^2(a_k, A\cap(B-v)) - \operatorname{dist}^2(a_{k+1}, A\cap(B-v)).\end{aligned}\tag{4.15}$$

Thus

$$\operatorname{dist}(a_{k+1}, A\cap(B-v))^2 \leq \operatorname{dist}(a_k, A\cap(B-v))^2 - \frac{1}{c_0^{2r}}\operatorname{dist}(a_k, A\cap(B-v))^{2r}.$$

Now, let $\beta_k := \operatorname{dist}(a_k, A\cap(B-v))^2$, $k\in\mathbb{N}$. Then, we have

$$\beta_{k+1} \leq \beta_k(1 - \frac{1}{c_0^{2r}}\beta_k^{r-1}).$$

Applying the preceding Lemma 4.1 with $\delta_k := \frac{1}{c_0^{2r}}$ and $p := r-1$, we see that

$$\operatorname{dist}^2(a_k, A\cap(B-v)) = \beta_k \leq \left(\beta_0^{1-r} + \frac{(r-1)}{c_0^{2r}}k\right)^{-\frac{1}{r-1}} \quad \text{for all} \quad k\in\mathbb{N}.$$

Thence there exists $M_0 > 0$ such that

$$\operatorname{dist}(a_k, A\cap(B-v)) \leq M_0\frac{1}{k^\rho}, \quad \forall k\in\mathbb{N}$$

where $\rho := \frac{1}{\min\left\{(2d-1)^n-1,\, 2\beta(n-1)d^n-2\right\}}$. Then, [6, Example 3.2] shows that $(a_k)_{k\in\mathbb{N}}$ is Fejér monotone with respect to $A\cap(B-v)$. Thus, by Fact 2.17,

$$\|a_k - \widetilde{a}\| \leq 2\operatorname{dist}(a_k, A\cap(B-v))^2 \leq 2M_0\frac{1}{k^\rho}, \quad \forall k\in\mathbb{N}.$$

Similarly, we can show that there exists $M_1 > 0$ such that

$$\operatorname{dist}(b_k, B\cap(A+v)) \leq M_1\frac{1}{k^\rho} \quad \text{and} \quad \|b_k - \widetilde{b}\| \leq 2M_1\frac{1}{k^\rho}, \quad \forall k\in\mathbb{N}.$$

Therefore, the conclusion follows by taking $M := \max\{2M_0, 2M_1\}$. □

5 Examples and remarks

In this section, we will provide several examples of the rates of convergence of the cyclic projection algorithm and the von Neumann alternating projection algorithm.

Example 5.1 Let

$$\begin{aligned} C_1 &:= \{(x,y) \in \mathbb{R}^2 \mid (x+1)^2 + y^2 - 1 \le 0\} \\ C_2 &:= \{(x,y) \in \mathbb{R}^2 \mid x + y - 1 \le 0\} \\ C_3 &:= \{(x,y) \in \mathbb{R}^2 \mid (x-1)^2 + y^2 - 1 \le 0\} \\ C_4 &:= \{(x,y) \in \mathbb{R}^2 \mid x + (y+2)^2 - 4 \le 0\}. \end{aligned}$$

Take $x_0 \in \mathbb{R}^2$. Let $(x_k)_{k\in\mathbb{N}}$ be defined by

$$x_1 := P_1x_0,\ x_2 := P_2x_1,\ x_3 := P_3x_2,\ x_4 := P_4x_3,\ x_5 := P_1x_4 \ldots$$

Then $\|x_k\| = O(\frac{1}{k^{\frac{1}{6}}})$. $\Diamond$

Proof. Clearly, $\bigcap_{i=1}^4 C_i = \{0\}$. Then apply $n = 2$ and $d = 2$ to Theorem 4.2. $\square$

Example 5.2 Let

$$\begin{aligned} A &:= \big\{(x,y) \in \mathbb{R}^2 \mid (x+1)^2 + y^2 - 1 \le 0\big\} \\ B &:= \big\{(x,y) \in \mathbb{R}^2 \mid (x-2)^2 + y^2 - 1 \le 0\big\}. \end{aligned}$$

Let $(a_k)_{k\in\mathbb{N}}, (b_k)_{k\in\mathbb{N}}$ be defined by

$$b_0 \in \mathbb{R}^2, \quad a_{k+1} := P_Ab_k, \quad b_{k+1} := P_Ba_{k+1}.$$

Then $\|a_k\| = O(\frac{1}{k^{\frac{1}{6}}})$, $\|b_k - (1,0)\| = O(\frac{1}{k^{\frac{1}{6}}})$. $\Diamond$

Proof. By the assumption, there exist unique points $a_0 \in \operatorname{bd} A$, $b_0 \in \operatorname{bd} B$ such that $1 = \operatorname{dist}(A,B) = \|a_0 - b_0\|$. Clearly, $a_0 = (0,0)$ and $b_0 = (1,0)$. Then, the conclusion follows by applying Theorem 4.7 with $n = 2$ and $d = 2$. $\square$

Example 5.3 Let $\alpha \ge 0$ and

$$\begin{aligned} A &:= \big\{(x,y) \in \mathbb{R}^2 \mid (x+1)^2 + y^2 - 1 \le 0\big\} = (-1,0) + \overline{\mathbb{B}}(0,1) \\ B &:= \big\{(x,y) \in \mathbb{R}^2 \mid -x + \alpha \le 0\big\}. \end{aligned}$$

Let $(a_k)_{k\in\mathbb{N}} := (u_k, v_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}} := (s_k, t_k)_{k\in\mathbb{N}}$ be defined by

$$b_0 \in \mathbb{R}^2, \quad a_{k+1} := P_Ab_k, \quad b_{k+1} := P_Ba_{k+1}.$$

Then for every $k \ge 2$

$$b_k = \Big(\alpha, \frac{t_1}{\sqrt{(1+\alpha)^{2(k-1)} + t_1^2 \sum_{i=0}^{k-2}((1+\alpha)^{2i}}}\Big)$$

$$a_{k+1} = \Big(-1 + \frac{\alpha+1}{\sqrt{(\alpha+1)^2 + \frac{t_1^2}{(1+\alpha)^{2(k-1)} + t_1^2\sum_{i=0}^{k-2}((1+\alpha)^{2i}}}}, \frac{t_1}{\sqrt{(1+\alpha)^{2k} + t_1^2 \sum_{i=0}^{k-1}((1+\alpha)^{2i}}}\Big).$$

Consequently, $a_k \longrightarrow 0$ and $b_k \longrightarrow (\alpha, 0)$ at the rate of $k^{-\frac{1}{2}}$ when $\alpha = 0$. When $\alpha \neq 0$(then $A \cap B = \varnothing$), $a_k \longrightarrow 0$ and $b_k \longrightarrow (\alpha, 0)$ at the rate of $(1+\alpha)^{-k}$. $\Diamond$

Proof. We first claim that

$$b_{k+1} = \big(\alpha, t_{k+1}\big) = \big(\alpha, \frac{t_k}{\sqrt{(1+\alpha)^2 + t_k^2}}\big), \quad \forall k \geq 1. \tag{5.1}$$

By [10, Examples 3.17&3.21 and Proposition 3.17], we have

$$P_A(x, y) = (-1, 0) + \frac{(x+1, y)}{\max\big\{1, \|(x+1, y)\|\big\}}, \quad \forall (x, y) \in \mathbb{R}^2 \tag{5.2}$$

$$P_B(x, y) = (\alpha, y), \quad \forall (x, y) \notin \operatorname{int} B.$$

Let $k \geq 1$. Since $A \cap B = \{0\}$ or $A \cap B = \varnothing$, $a_k \notin \operatorname{int} B$. Then by (5.2), $b_k = (\alpha, v_k)$ and then

$$a_{k+1} = P_A b_k = (-1, 0) + \frac{(1+\alpha, v_k)}{\max\big\{1, \|(1+\alpha, v_k)\|\big\}} = (-1, 0) + \frac{(1+\alpha, v_k)}{\sqrt{(1+\alpha)^2 + v_k^2}}$$

$$b_{k+1} = P_B(a_{k+1}) = (\alpha, \frac{v_k}{\sqrt{(1+\alpha)^2 + v_k^2}}) = (\alpha, \frac{t_k}{\sqrt{(1+\alpha)^2 + t_k^2}}).$$

Hence (5.1) holds. Next we show that

$$b_k = \Big(\alpha, \frac{t_1}{\sqrt{(1+\alpha)^{2(k-1)} + t_1^2 \sum_{i=0}^{k-2}((1+\alpha)^{2i}}}\Big), \quad \forall k \geq 2. \tag{5.3}$$

We prove (5.3) by the induction on k.

By (5.1), (5.3) holds when $k = 2$. Now assume that (5.3) holds when $k = p$, where $p \geq 2$. Now we consider the case of $k = p + 1$. By the assumption, we have

$$b_p = \Big(\alpha, \frac{t_1}{\sqrt{(1+\alpha)^{2(p-1)} + t_1^2 \sum_{i=0}^{p-2}((1+\alpha)^{2i}}}\Big). \tag{5.4}$$

Then by (5.1), we have

$$\begin{aligned} b_{p+1} &= \Big(\alpha, \frac{t_p}{\sqrt{(1+\alpha)^2 + t_p^2}}\Big) \\ &= \Big(\alpha, \frac{\frac{t_1}{\sqrt{(1+\alpha)^{2(p-1)} + t_1^2 \sum_{i=0}^{p-2}((1+\alpha)^{2i}}}}{\sqrt{(1+\alpha)^2 + \frac{t_1^2}{(1+\alpha)^{2(p-1)} + t_1^2 \sum_{i=0}^{p-2}(1+\alpha)^{2i}}}}\Big) \\ &= \Big(\alpha, \frac{t_1}{\sqrt{(1+\alpha)^{2p} + t_1^2 \sum_{i=0}^{p-1}((1+\alpha)^{2i}}}\Big). \end{aligned}$$

Hence (5.3) holds.

Combining(5.2) and (5.3), we have for every $k \geq 2$

$$
\begin{aligned}
a_{k+1} &= P_A b_k \\
&= \Big(-1 + \frac{\alpha+1}{\sqrt{(\alpha+1)^2 + \frac{t_1^2}{(1+\alpha)^{2(k-1)}+t_1^2\sum_{i=0}^{k-2}((1+\alpha)^{2i}}}}, \frac{t_1}{\sqrt{(1+\alpha)^{2k}+t_1^2\sum_{i=0}^{k-1}((1+\alpha)^{2i}}} \Big).
\end{aligned}
$$

Hence $a_k \longrightarrow 0$ and $b_k \longrightarrow (\alpha, 0)$ at the rate of $k^{-\frac{1}{2}}$ when $\alpha = 0$. When $\alpha \neq 0$, $a_k \longrightarrow 0$ and $b_k \longrightarrow (\alpha, 0)$ at the rate of $(1+\alpha)^{-k}$. $\square$

Remark 5.4 According to Theorem 4.7,we can only deduce that $(a_k)_{k\in\mathbb{N}}$ in Example 5.3 converges to $(0,0)$ and $(b_k)_{k\in\mathbb{N}}$ converge to $(\alpha, 0)$ at the rate of at least of $k^{-\frac{1}{6}}$. $\diamondsuit$

Example 5.5 Let

$$
\begin{aligned}
A &:= \big\{(x,y) \in \mathbb{R}^2 \mid (x+1)^2 + y^2 - 1 \leq 0\big\} \\
B &:= \big\{(x,y) \in \mathbb{R}^2 \mid (x-1)^2 + y^2 - 1 \leq 0\big\}.
\end{aligned}
$$

Let $(x_k, y_k)_{k\in\mathbb{N}}$ be defined by

$$(x_0, y_0) \in \mathbb{R}^2, (x_1, y_1) := P_A(x_0, y_0), (x_2, y_2) := P_B(x_1, y_1), (x_3, y_3) := P_A(x_2, y_2), \quad \cdots .$$

Note that

$$
\begin{aligned}
P_A(x,y) &= \left(-1 + \frac{x-1}{\sqrt{(x+1)^2+y^2}}, \frac{y}{\sqrt{(x+1)^2+y^2}}\right), \quad \forall (x,y) \in \mathbb{R}_+ \times \mathbb{R}_{++} \\
P_B(x,y) &= \left(1 + \frac{x+1}{\sqrt{(x-1)^2+y^2}}, \frac{y}{\sqrt{(x-1)^2+y^2}}\right), \quad \forall (x,y) \in \mathbb{R}_- \times \mathbb{R}_{++}.
\end{aligned}
$$

Figure 1 depicts the algorithm's trajectory with starting point $(0, 2)$.

Suppose, without loss of generality, that one starts on a point on one the half-circles nearest the other circle. Then the distance from zero (for every $k \in \mathbb{N}$), $r_k := \sqrt{x_k^2 + y_k^2}$ satisfies $r_k^2 = 2\alpha_k$ where $\alpha_k := |x_k|$ since $(x_k, y_k) \in \operatorname{bd} A \cup \operatorname{bd} B$. Hence

$$1 - \alpha_{k+1} = \frac{1+\alpha_k}{\sqrt{1+4\alpha_k}}.$$

Linearizing, we obtain that $w_k := 4\alpha_k$ approximately satisfies the logistics equation

$$w_{k+1} \approx w_k(1 - w_k)$$

This can be explicitly solved by writing

$$\frac{1}{w_{k+1}} - \frac{1}{w_k} = \frac{1}{1-w_k}$$

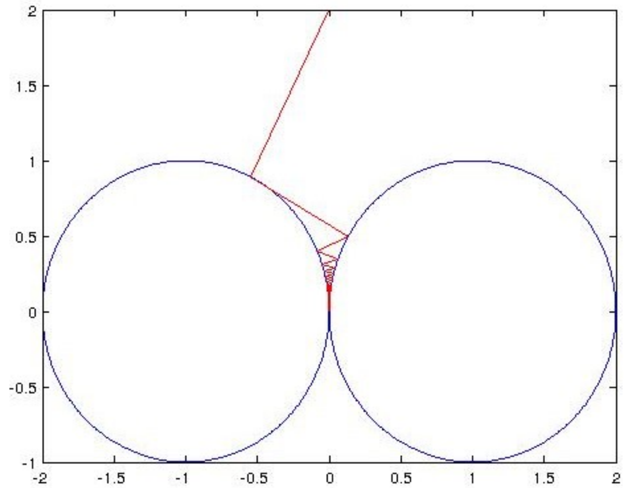

Figure 1: The iteration commencing at $(0, 2)$.

When summing and dividing by N, leads to

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{Nw_N} &= \lim_{N\to\infty} \left(\frac{1}{Nw_N} - \frac{1}{Nw_0}\right) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \frac{1}{1-w_k} \\
&= \lim_{N\to\infty} \frac{1}{1-w_N} = 1,
\end{aligned} \tag{5.5}$$

since Césaro summability is conservative and $w_N \to 0$. Hence $\alpha_k \sim 1/(4k)$ and so

$$\sqrt{x_k^2 + y_k^2} = r_k \sim \frac{1}{\sqrt{2k}}.$$

For instance, with $\alpha_0 = 1, N = 10^6$, we obtain $\alpha_N \approx 0.0000002499992442$. A similar analysis can be performed in the previous example. $\Diamond$

Remark 5.6 According to Theorem 4.7, we can only deduce that $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ in Example 5.5 converge to $(0, 0)$ at the rate of at least of $k^{-\frac{1}{6}}$. $\Diamond$

Example 5.7 Let A, B be defined by

$$\begin{aligned}
A &:= \{(x, y) \in \mathbb{R}^2 \mid x \le 0\} \\
B &:= \{(x, y) \in \mathbb{R}^2 \mid y^2 - x \le 0\}.
\end{aligned}$$

Let $(a_k)_{k\in\mathbb{N}}$, $(b_k)_{k\in\mathbb{N}} := (x_k, y_k)_{k\in\mathbb{N}}$ be defined by

$$b_0 := (x_0, y_0) \in \mathbb{R} \times \mathbb{R}_+ \text{ with } \|b_0\| \le 1, \quad a_{k+1} := P_A b_k, \quad b_{k+1} := P_B a_{k+1}.$$

Then for every $k \in \mathbb{N}$

$$a_{k+1} = (0, y_k), \quad b_k = (y_k^2, y_k), \quad \text{and} \quad y_{k+1} \approx y_k - 2y_k^3.$$

In consequence, $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ converge to 0 at the rate of at least of $k^{-\frac{1}{2}}$.

Proof. Let $k \in \mathbb{N}$. Since $A \cap B = \{0\}$, $b_k \notin \operatorname{int} A$. Since $b_k \in \operatorname{bd} B$ and then $x_k = y_k^2$. Thus $a_{k+1} = P_A b_k = P_A(x_k, y_k) = (0, y_k)$. Then we have

$$b_{k+1} = (x_{k+1}, y_{k+1}) = (y_{k+1}^2, y_{k+1}) = P_B a_{k+1} = P_B(0, y_k).$$

Thus (y_{k+1}^2, y_{k+1}) is a minimizer of the function

$$y \mapsto \|(y^2, y) - (0, y_k)\|^2 = \|(y^2, y - y_k)\|^2 = y^4 + y^2 - 2yy_k + y_k^2.$$

Thus

$$4y_{k+1}^3 + 2y_{k+1} - 2y_k = 0.$$

Then we have

$$y_{k+1} = \frac{1}{6}\Big(54y_k + 6\sqrt{6 + 81y_k^2}\Big)^{\frac{1}{3}} - \Big(54y_k + 6\sqrt{6 + 81y_k^2}\Big)^{-\frac{1}{3}}.$$

Linearizing the above equation, we have

$$y_{k+1} \approx y_k - 2y_k^3 = y_k(1 - 2y_k^2). \tag{5.6}$$

By (5.6), we have

$$y_{k+1}^2 \approx (y_k - 2y_k^3)^2 \approx y_k^2 - 4y_k^4 = y_k^2(1 - 4y_k^2).$$

Set $\omega_k := y_k^2$. Then we have

$$\omega_{k+1} \approx \omega_k(1 - 4\omega_k).$$

Similar to the corresponding lines in Example 5.5, we have $y_k \longrightarrow 0$ at the rate of exactly $k^{-\frac{1}{2}}$. Thence $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ converge to 0 at that rate. $\square$

Remark 5.8 Similarly, according to Theorem 4.7, we can only deduce that $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ in Example 5.7 converge to $(0, 0)$ at the rate of at least of $k^{-\frac{1}{6}}$. Using the technique in the proof of Example 5.7, we can show the sequence of the alternating projections in Example 5.3 converges to 0 at the rate of $k^{-\frac{1}{2}}$ when $\alpha = 0$. $\Diamond$

6 Conclusion and Open Questions

In this paper, we studied the rate of convergence of the cyclic projection algorithm applied to finitely many semi-algebraic convex sets. We established an explicit convergence rate estimate which relies on the maximum degree of the polynomials that generate the semi-algebraic convex sets and the dimension of the underlying space. We also examined some concrete examples and compared the actual convergence rate with our estimate.

Our results have suggested the following future research topics and open questions:

- The explicit examples (Example 5.3, 5.5 and 5.7) show that, in general, our estimate of the convergence rate of the cyclic projection algorithm will not be tight. It would be interesting to see how one can sharpen the estimate obtained in this paper and get a tight estimate for the cyclic projection algorithm. In particular, finding the right exponent when each set is defined by convex quadratic functions would be a good starting point.
- Can we extend the approach here to analyze the convergence rate of the Douglas-Rachford algorithm?

These will be our future research topics and will be examined later on.

Acknowledgments

Jonathan Borwein, Guoyin Li and Liangjin Yao were partially supported by various Australian Research Council grants.

References

[1] A. Auslender and M. Teboulle, *Asymptotic cones and functions in optimization and variational inequalities*. Springer Monographs in Mathematics. Springer-Verlag, New York, 2003.

[2] A. Auslender, Existence of optimal solutions and duality results under weak conditions, *Mathematical Programming* Vol. 88, pp 45-59

[3] J.B. Baillon, R.E Bruck, and S, Reich, On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces, *Houston Journal of Mathematics*, vol. 4, pp. 1–9, 1978.

[4] J.-B. Baillon, P. L. Combettes, and R. Cominetti, There is no variational characterization of the cycles in the method of periodic projections, *Journal of Functional Analysis*, vol. 262, pp. 400–408, 2012.

[5] H.H. Bauschke, Projection algorithms: results and open problems, *Inherently parallel algorithms in feasibility and optimization and their applications*, pp. 11-22, 2001.

[6] H.H. Bauschke and J.M. Borwein, On the convergence of von Neumanns alternating projection algorithm for two sets, *Set-Valued Analysis*, vol. 1, pp. 185–212, 1993.

[7] H.H. Bauschke and J.M. Borwein, Dykstra's alternating projection algorithm for two sets, *Journal of Approximation Theory* , vol. 79, pp. 418–443, 1994.

[8] H.H. Bauschke and J.M. Borwein: On projection algorithms for solving convex feasibility problems, *SIAM Review*, vol. 38, pp. 367-426, 1996.

[9] H.H. Bauschke, J.M. Borwein, and A.S. Lewis, The method of cyclic projections for closed convex sets in Hilbert space, *Recent developments in optimization theory and nonlinear analysis*, Contemporary Mathematic, pp. 1–38, 1997

[10] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.

[11] H.H. Bauschke, D.R. Luke, H. M. Phan, and X. Wang, Restricted normal cones and the method of alternating projections, *Set-Valued and Variational Analysis*, in press, `http://arxiv.org/abs/1205.0318v1`.

[12] E.G. Belousov, *Introduction to convex analysis and integer programming*, Izdat. Moskov. Univ., Moscow, 1977

[13] E.G. Belousov, On types of Hausdorff discontinuity from above for convex closed mappings, *Optimization*, vol. 49, pp. 303–325, 2001.

[14] J.M. Borwein and J.D. Vanderwerff, *Convex Functions*, Cambridge University Press, 2010.

[15] R.E Bruck, and S, Reich, Nonexpansive projections and resolvents of accretive operators in Banach spaces, *Houston Journal of Mathematics*, vol. 3, pp. 459–470, 1977.

[16] E. G. Belousov and D. Klatte, A Frank-Wolfe type theorem for convex polynomial programs, Computational Optimization and Applications, vol. 22, pp. 37-48, 2002.

[17] J. Bochnak, M. Coste and M.F. Roy, *Real algebraic geometry*, Springer-Verlag, Berlin, 1998.

[18] L.M. Brègman, Finding the common point of convex sets by the method of successive projection. *Doklady Akademii Nauk SSSR* , vol. 162, pp. 487–490, 1965.

[19] R.S. Burachik and A.N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer-Verlag, 2008.

[20] D. Butnariu and A.N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, Kluwer Academic Publishers, 2000.

[21] S. Deng, Well-posed problems and error bounds in optimization. Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997), Appl. Optim., (22) 1999 Kluwer Acad. Publ., Dordrecht, 117-126.

[22] F. Deutsch, *Best approximation in inner product spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 7, Springer-Verlag, New York, 2001.

[23] F. Deutsch and H. Hundal, The rate of convergence for the cyclic projections algorithm. I. Angles between convex sets, *Journal of Approximation Theory*, vol. 142, pp. 3655, 2006.

[24] F. Deutsch and H. Hundal, The rate of convergence for the cyclic projections algorithm. II. Norms of nonlinear operators, summary, *Journal of Approximation Theory*, vol. 142, pp. 56–82, 2006.

[25] J. Gwoździewicz, The Łojasiewicz exponent of an analytic function at an isolated zero, Commentarii Mathematici Helvetici, vo. 74, pp. 364–375, 1999.

[26] L.G. Gubin; B.T. Polyak, and E.V. Raik, The method of projections for finding the common point of convex sets, *USSR Computational Mathematics and Mathematical Physics*, vol. 7, pp. 1–24, 1967.

[27] J.B. Hiriart-Urruty and C. Lemarachal, Convex analysis and minimization algorithms. II. Advanced theory and bundle methods. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 306. Springer-Verlag, Berlin, 1993.

[28] D. Klatte, Hoffman's error bound for systems of convex inequalities. Mathematical programming with data perturbations, 185–199, Lecture Notes in Pure and Appl. Math., Dekker, New York, 1998.

[29] J. Kollár, An effective Łojasiewicz inequality for real polynomials, *Periodica Mathematica Hungarica*, vol. 38, pp. 213–221, 1999.

[30] B. Kummer, Stability and weak duality in convex programming without regularity. Prepint. Humboldt Univ. Berlin, 1978.

[31] A.S. Lewis and J. S. Pang, Error bounds for convex inequality systems Generalized Convexity, Generalized Monotonicity J.P. Crouzeix, J.E. Martinez-Legaz and M. Volle (eds), pp. 75–110, 1998.

[32] G. Li, On the asymptotic well behaved functions and global error bound for convex polynomials, *SIAM Journal on Optimization*, vol. 20, pp. 1923–1943, 2010.

[33] G. Li and K.F. Ng, Error bounds of generalized D-gap functions for nonsmooth and nonmonotone variational inequality problems, *SIAM Journal on Optimization*, vol. 20, pp. 667–690, 2009.

[34] G. Li and B.S. Mordukhovich, Hölder metric subregularity with applications to proximal point method, *SIAM Journal on Optimization*, vol. 22, pp. 1655–1684, 2012.

[35] W. Li, Error bounds for piecewise convex quadratic programs and applications, *SIAM J. on Control and Optim.*, vol. 33, pp. 1510– 1529, 1995.

[36] X.D. Luo and Z.Q. Luo, Extension of Hoffman's error bound to polynomial systems, *SIAM Journal on Optimization*, vol. 4,pp. 383–392, 1994.

[37] Z.Q. Luo and J.S. Pang, Error bounds for analytic systems and their applications, *Mathematical Programming*, vol. 67, pp. 1–28, 1994.

[38] Z.Q. Luo, J.S. Pang, and D. Ralph, Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge, 1996.

[39] W.T. Obuchowska, On generalizations of the Frank-Wolfe theorem to convex and quasi-convex programmes, Computational Optimization and Applications, vol. 33, pp. 349–364, 2006.

[40] J.S. Pang, Error bounds in mathematical programming. *Mathematical Programming*, vol. 79, pp. 299-332, 1997.

[41] R.R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, 2nd Edition, Springer-Verlag, 1993.

[42] R.T. Rockafellar and R.J-B Wets, *Variational Analysis*, 3nd Printing, Springer-Verlag, 2009.

[43] V.M Shironin, *On Hausdorff continuity of convex and convex polynomial mappings*, in Mathematical Optimization: Questions of Solvability and Stability, E.G. Belousov and B. Bank, eds., Moscow University Publishing, Moscow, 1986.

[44] von Neumann, *Functional Operators, Vol.II*, Princeton University Press, 1950.

[45] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing, 2002.